

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 13:04:25 ; Search time 935 Seconds
(without alignments)
12240.273 Million cell updates/sec

Title: US-09-497-822C-18

Perfect score: 5082

Sequence: 1 gagctctgacaaattgag.....acaagcaaaacaaaaaaa 5082

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	4188.2	82.4	4321	24	Human androgen rec
2	3595.8	70.8	3715	20	Human androgen rec
3	3595.8	70.8	3715	24	Gene #2300 used to
4	3594.2	70.7	3715	18	Full-length human
5	3547	69.8	3569	18	Androgen receptor
6	3545.4	69.8	3569	23	Human androgen rec
7	3498.2	68.8	3590	23	DNA encoding novel
8	2486.4	48.9	4180	10	Rat androgen recep
9	2244.2	44.2	3715	24	Human DNA for stag

c	10	2244.2	44.2	3715	24	ABK31545	Signal transductio
	11	2189.2	43.1	3217	12	AAQ12002	Full-length rat an
	12	2185.8	43.0	3217	10	AAK91578	Rat androgen recep
	13	2154	42.4	3715	24	ABK34012	Human DNA for stag
	14	2154	42.4	3715	24	ABK31544	Signal transductio
	15	1715.6	33.8	1810	22	AAF84342	Human androgen rec
c	16	1587	31.2	1998	23	AAK91973	DNA encoding novel
	17	1532.6	30.2	6905	24	ABK29886	Androgen receptor
	18	1340	26.4	2517	24	ABA01683	Fused androgen rec
	19	1148	22.6	1329	24	ABN85656	Human androgen rec
	20	834.4	16.4	1171	24	ABN85657	Human androgen rec
c	21	761.6	15.0	1432	24	ABL34259	Human immune syste
	22	711.8	14.0	1893	12	AAQ12008	Human immune syste
	23	703.6	13.8	1731	12	AAQ12007	TrpE/N-terminal do
	24	703.2	13.8	1432	24	ABL34258	Human immune syste
	25	611.6	12.0	612	21	AAK70311	Single nucleotide
	26	611.6	12.0	612	21	AAK70335	Single nucleotide
	27	575.6	11.3	598	21	AAK70305	Single nucleotide
	28	575.6	11.3	598	21	AAK70308	Single nucleotide
	29	524.2	10.3	5168	24	ABK94716	Neurodegenerative
	30	401.6	7.9	6244	16	AAQ76269	PhCMV*-1-controlle
	31	401.6	7.9	6244	20	AAK81724	cDNA encoding rabb
	32	401.6	7.9	6244	20	AAK27905	Rabbit progesteron
	33	401.6	7.9	6244	20	AAK01369	PhCMV*-1 rabbit pr
	34	401.6	7.9	6244	22	AAK09837	Rabbit progesteron
	35	400	7.9	6244	16	AAK06872	PhCMV*-1 promoter
	36	384	7.6	1231	14	AAQ37761	Sequence of a 1.23
	37	384	7.6	3014	18	AAK94756	Human progesterone
	38	384	7.6	3014	18	AAK84747	Human progesterone
	39	383	7.5	2970	21	AAK53851	Human progesterone
	40	343.2	6.8	360	22	AAK19609	Human breast cance
	41	340.6	6.7	1329	12	AAQ12009	TrpE/AR androgen-b
	42	320.8	6.3	358	21	AAK08780	Human secreted pro
	43	320.8	6.1	6322	24	ABK63639	vat sequence diffe
	44	310.8	6.1	7257	18	AAK68656	Green fluorescent
	45	305.6	6.0	5749	24	ABL61813	Colon adenocarcino

ALIGNMENTS

RESULT 1

AAK30440
ID AAK30440 standard; DNA; 4321 BP.

XX AC AAD30440;

XX DT 21-MAY-2002 (first entry)

XX DE Human androgen receptor (AR) gene.

XX KW Human; AIB1; amplified in breast cancer 1; androgen receptor; AR;
KW KW prostate cancer; chromosome X; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 1115..3874

XX FT /*tag= a

XX FT /product= "Human androgen receptor (AR) protein"

XX PN WO200210452-A2.

XX PD 07-FEB-2002.

XX PF 27-JUL-2001; 2001WO-US23834.

XX PR 27-JUL-2000; 2000US-221074P.

XX XX (UYRP) UNIV ROCHESTER.

XX PA Chang C;

XX PI

XX XX

DR WPI: 2002-206195/26.
DR P-PSDB; AAE19061.
XX
PT Assessing the risk of acquiring or developing prostate cancer in a
PT human subject, comprises determining the length of the contiguous CAG,
PT CAA and/or GGN repeats in the AIB1 gene and/or androgen receptor gene
PT of the subject -
XX
PS Disclosure; Page 61-62; 86pp; English.
XX
CC The invention relates to a method for assessing the risk of prostate
CC cancer in a human subject. The method involves determining the length
CC of the contiguous CAG or CAA repeats in both AIB1 (Amplified In Breast
CC cancer 1) gene alleles or contiguous CAG, CAA or GGN repeats in the
CC androgen receptor gene of the subject. The method is useful for
CC assessing a subject's risk for acquiring or developing prostate cancer.
CC The present sequence is human androgen receptor (AR) gene. Human
CC AR gene is located on X chromosome.
XX
SQ Sequence 4321 BP; 966 A; 1281 C; 1168 G; 906 T; 0 other;
Query Match 82.4%; Score 4188.2; DB 24; Length 4321;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 4297; Conservative 0; Mismatches 16; Indels 24; Gaps 8;
QY 754 CGAGATCCCGGGGAGCCAGCTTCTGGGAGAGCGGGAAGCGTCCGGAGCAAGCCACAGAG 813
Db 1 CGAGATCCCGGGGAGCCAGCTTCTGGGAGAGCGGGAAGCGTCCGGAGCAAGCCACAGAG 59
QY 814 CAGAGGAGGCGCAGAGGAGGAAAGGCGCCNAGCTAGCGCTCCAGTGTGTACAGNAGC 873
Db 60 CAGAGGAGGCGCAGAGGAGGAAAGGCGCCNAGCTAGCGCTCCAGTGTGTACAGNAGC 118
QY 874 CGAA-GGAGCCACAGCCAGCCAGCCAGCCGCTCCAGCAGAGCAAGCCCTTTTGA- 931
Db 119 CGAAGGAGCAGCAGCCAGCCAGCCAGCCGCTCCAGCAGAGCAAGCCCTTTTGA- 178
QY 932 -----NCGGTTGCAAGCGCGCGCGCGAGCTGCCCTTTCTCTTCGGTGAAGTTTAA 986
Db 179 CGCGCGCGCTCGAAGCGCGCGCGCGAGCTGCCCTTTCTCTTCGGTGAAGTTTAA 238
QY 987 AGCTGCTAAAGACTCGGAGGAGCAAGGAAGTGCCTGTAGACACTGAGCGCTGCTTTG 1046
Db 239 AGCTGCTAAAGACTCGGAGGAGCAAGGAAGTGCCTGTAGACACTGAGCGCTGCTTTG 298
QY 1047 TCCTCTCTCTCCAGCCCGCTCCCGCCAGCCCTGCTTCCCGCCCTCCCGCTCTCTC 1106
Db 299 TCCTCTCTCTCCAGCCCGCTCCCGCCAGCCCTGCTTCCCGCCCTCCCGCTCTCTC 358
QY 1107 TCCCGCAGCTGCTCAGTGGCTACTCTAGCCCAAGCCCGCTCAGCAGCCCTTCTCCCG 1166
Db 359 TCCCGCAGCTGCTCAGTGGCTACTCTAGCCCAAGCCCGCTCAGCAGCCCTTCTCCCG 418
QY 1167 CCGCCCGCCCGCCCGCTCGGCGCCAGCGTGNAGNCCGAGTTTGCAGAGGTAACCTCC 1226
Db 419 CCGCCCGCCCGCCCGCTCGGCGCCAGCGTGCAGCCCGAGTTTGCAGAGGTAACCTCC 478
QY 1227 CTTTGGCTGCGAGCGGCGAGNCTAGTGCACATGCAAGAGGCTCTTAGAG-CAGG 1285
Db 479 CTTTGGCTGCGAGCGGCGAG-CTAGCTGCACATGCAAGAGGCTCTTAGAGCGCAGG 537
QY 1286 CGACTGGGAGCGGCTCAGCACTGACGACGACGACGCGCTGGTTAGGCTGACGCGGAG 1345
Db 538 CGACTGGGAGCGGCTCAGCACTGACGACGACGACGCGCTGGTTAGGCTGACGCGGAG 597
QY 1346 AGAACCTCTGTTTCCCGCCACTCTCTCTCCACCTCTCTCTGCTTCCCGCCCGAGTG 1405
Db 598 AGAACCTCTGTTTCCCGCCACTCTCTCTCCACCTCTCTCTGCTTCCCGCCCGAGTG 657
QY 1406 CGGAGCCAGAGATCAAAAGATGAAAGGAGTCAAGTCTCAGTGTAGCCAAACAAAC 1465
Db 658 CGGAG-CAGAGATCAAAAGATGAAAGGAGTCAAGTCTCAGTGTAGCCAAACAAAC 716

QY 1466 AAACAAAACAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1525
Db 717 AAACAAAACAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 776
QY 1526 CTACTTCAGTGGACACTGAATTTGGAGGTGGAGGATTTGTTTCTTTTCTTTTAAATCT 1585
Db 777 CTACTTCAGTGGACACTGAATTTGGAGGTGGAGGATTTGTTTCTTTTCTTTTAAATCT 836
QY 1586 GGGCATCTTTTGAATCTTACCTTTCAAGTATTAAGAGACAGACTGTGAGCCTAGCAGGGCA 1645
Db 837 GGGCATCTTTTGAATCTTACCTTTCAAGTATTAAGAGACAGACTGTGAGCCTAGCAGGGCA 896
QY 1646 GATCTTCACCGCTGTGTCTTCTTCGACAGAGACTTTGAGCTGTGAGAGGCTTTT 1705
Db 897 GATCTTCACCGCTGTGTCTTCTTCGACAGAGACTTTGAGCTGTGAGAGGCTTTT 956
QY 1706 GCGTGGTGTCTCCCGCAAGTTTCTTCTCTGAGCTTCCCGCAGGTGGGAGCTAGCTGC 1765
Db 957 GCGTGGTGTCTCCCGCAAGTTTCTTCTCTGAGCTTCCCGCAGGTGGGAGCTAGCTGC 1016
QY 1766 AGCGACTACCGCATCATCACAGCTGTGAACTCTTCTGAGCAAGAGAGGAGGCGGG 1825
Db 1017 AGCGACTACCGCATCATCACAGCTGTGAACTCTTCTGAGCAAGAGAGGAGGCGGG 1076
QY 1826 GTAAGGGAAGTAGTGAAGATTCAGCAAGCTCAAGATGGAAGTCACTTAGGGCTGG 1885
Db 1077 GTAAGGGAAGTAGTGAAGATTCAGCAAGCTCAAGATGGAAGTCACTTAGGGCTGG 1136
QY 1886 GAAGGCTTACCTTCGCGCGCGCTCCAGAGCTTACCAGAGAGCTTCCAGAACTGTGTCC 1945
Db 1137 GAAGGCTTACCTTCGCGCGCGCTCCAGAGCTTACCAGAGAGCTTCCAGAACTGTGTCC 1196
QY 1946 AGAGGCTGCGGAAGTGAATCCAGAACCCGGGCCCGCCAGCCAGAGGCGCGGAGCCGAG 2005
Db 1197 AGAGGCTGCGGAAGTGAATCCAGAACCCGGGCCCGCCAGCCAGAGGCGCGGAGCCGAG 1256
QY 2006 CACCTCCCGCGCGCGCTTGTGTCTGTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2065
Db 1257 CACCTCCCGCGCGCGCTTGTGTCTGTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1304
QY 2066 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2125
Db 1305 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1364
QY 2126 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2185
Db 1365 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1424
QY 2186 GCTACTCTGCTCTGTGATGAGGAACAGCAACCTTTCACAGCCGAGTTCGCGCCCTGAGGTGCC 2245
Db 1425 GCTACTCTGCTCTGTGATGAGGAACAGCAACCTTTCACAGCCGAGTTCGCGCCCTGAGGTGCC 1484
QY 2246 ACCCGAGAGAGTTCGCTCCAGAGCCTGGAGCGCGCTGGCGCCAGCAAGGCGGTGC 2305
Db 1485 ACCCGAGAGAGTTCGCTCCAGAGCCTGGAGCGCGCTGGAGCGCGCTGGCGCCAGCAAGGCGGTGC 1544
QY 2306 CCGCAGCAGCTGCCAGCAGCTCCCGGACGAGGATGACTCAGCTGCCCGCATCCAGCTGTGCC 2365
Db 1545 CCGCAGCAGCTGCCAGCAGCTCCCGGACGAGGATGACTCAGCTGCCCGCATCCAGCTGTGCC 1604
QY 2366 TGTGTGGGCGCCACTTTCGCGGCTTAAAGCAGTGTCTCGCTGACCTTAAAGACATCTCTGA 2425
Db 1605 TGTGTGGGCGCCACTTTCGCGGCTTAAAGCAGTGTCTCGCTGACCTTAAAGACATCTCTGA 1664
QY 2426 GCGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2485
Db 1665 GCGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1724
QY 2486 GCGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2545
Db 1725 GCGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1784
QY 2546 TAGGGGCGACTTCGACCATTTTCTGACAGCCGAGGAGTGTGTAAAGGCAAGTGTCTGGTGT 2605

Db 1785 TAGGGGACATTCGACCATTTCTGACAAACCCAGAGATTGTGAAGGCAAGTGTGCGGTG 1844
QY 2606 CCATGGGCGTGGTGTGGAGCGGTGGAGCATCTGAGTCCAGGGGAACAGCTTCGGGGGG 2665
Db 1845 CCATGGGCGTGGTGTGGAGCGGTGGAGCATCTGAGTCCAGGGGAACAGCTTCGGGGGG 1904
QY 2666 ATTGCATGTACGCCCCACTTTTGGAGTTCCACCCGCTGTGCGTCCCACTCTCTGTGCC 2725
Db 1905 ATTGCATGTACGCCCCACTTTTGGAGTTCCACCCGCTGTGCGTCCCACTCTCTGTGCC 1964
QY 2726 CATTTGGCCGAATGCAAGGTTCTCTGTAGACAGACAGCGCAGCAGAGACACTGAAGATA 2785
Db 1965 CATTTGGCCGAATGCAAGGTTCTCTGTAGACAGACAGCGCAGCAGAGACACTGAAGATA 2024
QY 2786 CTGCTGAGTATTCCTCTTCAAGGAGGTTACACCAAGGGCTAGAGCGGAGAGCCTAG 2845
Db 2025 CTGCTGAGTATTCCTCTTCAAGGAGGTTACACCAAGGGCTAGAGCGGAGAGCCTAG 2084
QY 2846 GCTGCTCTGGCAGCGCTGACGAGGAGCTCCGGGACACTTGAACCTGCCGTCTACCCCTGT 2905
Db 2085 GCTGCTCTGGCAGCGCTGACGAGGAGCTCCGGGACACTTGAACCTGCCGTCTACCCCTGT 2144
QY 2906 CTCTCTACAGTCCGGAGCACTGGAGGAGCTGGGTACGAGTCCGCACTACTACA 2965
Db 2145 CTCTCTACAGTCCGGAGCACTGGAGGAGCTGGGTACGAGTCCGCACTACTACA 2204
QY 2966 ACTTTCACACTGGCTTGGCCGACCGCCGCCCTCCGCGCTCCGCGGCTCCGCGCGCAGT 3025
Db 2205 ACTTTCACACTGGCTTGGCCGACCGCCGCCCTCCGCGCTCCGCGGCTCCGCGCGCAGT 2264
QY 3026 GCATCAAGCTGGAGAACCGCTGGACTAGCGAGCGCTGGCGGCTGGCGGCGCGAGT 3085
Db 2265 GCATCAAGCTGGAGAACCGCTGGACTAGCGAGCGCTGGCGGCTGGCGGCGCGAGT 2324
QY 3086 GCGCTATGGGACCTGGCGAGCTTCCTCATCTCTGGACACTCTCTTTCACAGCCGAAGAGCCAGT 3145
Db 2325 GCGCTATGGGACCTGGCGAGCTTCCTCATCTCTGGACACTCTCTTTCACAGCCGAAGAGCCAGT 2384
QY 3146 CACCTCAGCCCGCTTCTCTATCTCTGGACACTCTCTTTCACAGCCGAAGAGCCAGT 3205
Db 2385 CACCTCAGCCCGCTTCTCTATCTCTGGACACTCTCTTTCACAGCCGAAGAGCCAGT 2444
QY 3206 TGTATGGACCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3265
Db 2445 TGTATGGACCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2504
QY 3266 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3325
Db 2505 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2564
QY 3326 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3385
Db 2565 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2624
QY 3386 CTGGGCGCATGGTGGAGCAGTGGCTTATCCAGTCCACTTGTGTCAAAAGGGAATGG 3445
Db 2625 CTGGGCGCATGGTGGAGCAGTGGCTTATCCAGTCCACTTGTGTCAAAAGGGAATGG 2684
QY 3446 GCCCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3505
Db 2685 GCCCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2744
QY 3506 ACCATGTTTGGCCATTTGACTTATTTCTTCCACCCAGAGACCTTCCTGATCTGTGGAG 3565
Db 2745 ACCATGTTTGGCCATTTGACTTATTTCTTCCACCCAGAGACCTTCCTGATCTGTGGAG 2804
QY 3566 ATGAAGCTTCTGGGTGTCACTATGGAGCTCTACATGTGAAGCTGCAAGGCTTCTTCA 3625
Db 2805 ATGAAGCTTCTGGGTGTCACTATGGAGCTCTACATGTGAAGCTGCAAGGCTTCTTCA 2864
QY 3626 AAAGAGCGCTGAAGGGAACAAAGTACCTGTGGCCAGAGAAATGATTCATATG 3685

Db 2865 AAAGAGCGCTGAAGGGAACAAAGTACCTGTGGCCAGCAGAAATGATTCATATTTG 2924
QY 3686 ATAAATTTCCAAAGAAATTTGCTCATTTGCTGCTTCCGAAATGTTATGAAGAGGGA 3745
Db 2925 ATAAATTTCCAAAGAAATTTGCTCATTTGCTGCTTCCGAAATGTTATGAAGAGGGA 2984
QY 3746 TGAATCTGGGAGCCCGGAGCTGAAGAACTTGGTAATCTGAACCTACAGGAGCAAGAG 3805
Db 2985 TGAATCTGGGAGCCCGGAGCTGAAGAACTTGGTAATCTGAACCTACAGGAGCAAGAG 3044
QY 3806 AGCTTCCAGCACCCAGCCCACTGAGGAGACAAACCCAGAGCTGACATGTGCACACA 3865
Db 3045 AGCTTCCAGCACCCAGCCCACTGAGGAGACAAACCCAGAGCTGACATGTGCACACA 3104
QY 3866 TTGAAGGCTATGAATGTACAGCCCACTTCTGAATGTCTTGAAGCCATTGAGCAGGTG 3925
Db 3105 TTGAAGGCTATGAATGTACAGCCCACTTCTGAATGTCTTGAAGCCATTGAGCAGGTG 3164
QY 3926 TAGTGTGTCTGGACAGACAAACACCCAGCCCACTTCTTGCAGCCCTTGCCTCTAGCC 3985
Db 3165 TAGTGTGTCTGGACAGACAAACACCCAGCCCACTTCTTGCAGCCCTTGCCTCTAGCC 3224
QY 3986 TCAATGAAGCTGGAGAGACAGCTTGTACAGTGGTCAAGTGGGCCAAGGCCCTTGCTG 4045
Db 3225 TCAATGAAGCTGGAGAGACAGCTTGTACAGTGGTCAAGTGGGCCAAGGCCCTTGCTG 3284
QY 4046 GCTTCCGCAACTTACACCTGGAGCAGATGGCTGTCTTCAATTCAGTACTCTCTGGAGTGG 4105
Db 3285 GCTTCCGCAACTTACACCTGGAGCAGATGGCTGTCTTCAATTCAGTACTCTCTGGAGTGG 3344
QY 4106 TCATGGTGTTCGCGATGGCTGGCGATCCTTTCACCAATGTCAACTCCAGGATGCTTACT 4165
Db 3345 TCATGGTGTTCGCGATGGCTGGCGATCCTTTCACCAATGTCAACTCCAGGATGCTTACT 3404
QY 4166 TCGCCCTCTGATCTGCTTTTCAATGAGTACCGCATGCGATGCGAAGTCCCGGATGTACAGCCAGT 4225
Db 3405 TCGCCCTCTGATCTGCTTTTCAATGAGTACCGCATGCGAAGTCCCGGATGTACAGCCAGT 3464
QY 4226 GTGTCCGAATGAGCCACTCTCTCAAGAGTTTGGATGGCTCCAAATCACCCCCAGGAAT 4285
Db 3465 GTGTCCGAATGAGCCACTCTCTCAAGAGTTTGGATGGCTCCAAATCACCCCCAGGAAT 3524
QY 4286 TCCTGTGATGAAAGCACTGCTACTCTTCAAGCATTATTTCCAGTGGATGGCTGAAATC 4345
Db 3525 TCCTGTGATGAAAGCACTGCTACTCTTCAAGCATTATTTCCAGTGGATGGCTGAAATC 3584
QY 4346 AAAATTTCTTGTGATGAACTTCGAATGAACATACATCAAGGAACCTCGATCGTATCATTTGCAT 4405
Db 3585 AAAATTTCTTGTGATGAACTTCGAATGAACATACATCAAGGAACCTCGATCGTATCATTTGCAT 3644
QY 4406 GCAAAAGAAAAATCCACATCTCTCAAGAGCTTCTTACCAGCTCACCAAGCTCCTGG 4465
Db 3645 GCAAAAGAAAAATCCACATCTCTCAAGAGCTTCTTACCAGCTCACCAAGCTCCTGG 3704
QY 4466 ACTCCGTGAGCCCTTATTCGAGAGAGTGCATCAGTTTGTGACCTTGTCTTCAAGT 4525
Db 3705 ACTCCGTGAGCCCTTATTCGAGAGAGTGCATCAGTTTGTGACCTTGTCTTCAAGT 3764
QY 4526 CACATGTTGAGCGTGGACTTTCCGGAATGATGCGAGATCATCTCTTGTGAAGTGC 4585
Db 3765 CACATGTTGAGCGTGGACTTTCCGGAATGATGCGAGATCATCTCTGTGCAAGTGC 3824
QY 4586 CCAAGATCTTCTTGGGAAAGTCAAGCCATCTATTTCACACCCAGTGAACATTTGAA 4645
Db 3825 CCAAGATCTTCTTGGGAAAGTCAAGCCATCTATTTCACACCCAGTGAACATTTGAA 3884
QY 4646 ACCCTATTTCCCAACCCAGCTCATGCCCTTTCAGATGTCTTGTGCTGTATTAAGTGC 4705
Db 3885 ACCCTATTTCCCAACCCAGCTCATGCCCTTTCAGATGTCTTGTGCTGTATTAAGTGC 3944
QY 4706 TGCATCTACTCTGTCAGTGGCTTGGGAAATTTCTCTATTGATGATGATGATGATGATGATG 4765
Db 3945 TGCATCTACTCTGTCAGTGGCTTGGGAAATTTCTCTATTGATGATGATGATGATGATGATG 4004

Db 759 TAGCCCCAGGACAGCAGCAGCAGCAGGCTGAGGATGTTCTCCCCAAGCCCATCGTAG 818
QY 2175 AGGCCCCACAGGCTACTGCTCTGATGATGAGAACAGACACCTTACAGCCGCGAGTCGGC 2234
Db 819 AGGCCCCACAGGCTACTGCTCTGATGAGGAACAGAACCTTACAGCCGCGAGTCGGC 878
QY 2235 CTTGGAGTGCACCCCGAGAGAGGTTGCTCCAGAGCCTGAGAGCCGCTGGCCGCCAG 2294
Db 879 CTTGGAGTGCACCCCGAGAGAGGTTGCTCCAGAGCCTGGAGCCGCTGGCCGCCAG 938
QY 2295 CAAGGGCTGCGCAGCAGCTGCCAGACCTCCCGACCTCCGACGAGTACTCAGTGCCTCCATC 2354
Db 939 CAAGGGCTGCGCAGCAGCTGCCAGACCTCCCGACCTCCGACGAGTACTCAGTGCCTCCATC 998
QY 2355 CAGTTGTCCTGCTGCTGGGCCACCTTCCCGGCTTAAGCAGCTGCTCCGCTGACCTTAA 2414
Db 999 CAGTTGTCCTGCTGCTGGGCCACCTTCCCGGCTTAAGCAGCTGCTCCGCTGACCTTAA 1058
QY 2415 AGACATCTGAGGAGGCGCAGCACCACCTCTTCCAGCAACAGCAGCAGGAAGCAGT 2474
Db 1059 AGACATCTGAGGAGGCGCAGCACCACCTCTTCCAGCAACAGCAGCAGGAAGCAGT 1118
QY 2475 ATCCGAAGCAGCAGCAGGAGAGGAGGAGGCTCGGGGCTCCGCTCCCTCCAA 2534
Db 1119 ATCCGAAGCAGCAGCAGGAGAGGAGGAGGCTCGGGGCTCCGCTCCCTCCAA 1178
QY 2535 GGACAATTACTTAGGGGCACTTCGACCATTTCTGACACGCCAAGAGTGTGTAAGGC 2594
Db 1179 GGACAATTACTTAGGGGCACTTCGACCATTTCTGACACGCCAAGAGTGTGTAAGGC 1238
QY 2595 AGTGTGCTGTCATGAGGCTGGGTGGAGGCTTCCAGGAGTTCACCCGCTGCTGCCAC 2714
Db 1239 AGTGTGCTGTCATGAGGCTGGGTGGAGGCTTCCAGGAGTTCACCCGCTGCTGCCAC 1358
QY 2715 TCCCTGTGCCCATTTGCCGAATGCAAGGTTCTCTGCTAGAGCAGCAGCGCAGGAAG 2774
Db 1359 TCCCTGTGCCCATTTGCCGAATGCAAGGTTCTCTGCTAGAGCAGCAGCGCAGGAAG 1418
QY 2775 CACTGAGATGCTGCTGAGTATTCCTCTTCAAGGAGGTTACACCAAGGCTACAGG 2834
Db 1419 CACTGAGATGCTGCTGAGTATTCCTCTTCAAGGAGGTTACACCAAGGCTACAGG 1478
QY 2835 CGAGAGCTAGGCTGCTGCGAGCGCTGCAGCAGGAGCTCCGGGACACTTGAACCTGCC 2894
Db 1479 CGAGAGCTAGGCTGCTGCGAGCGCTGCAGCAGGAGCTCCGGGACACTTGAACCTGCC 1538
QY 2895 GTCTACCTGCTCTCTACAACTCCGAGCCTGAGGAGGTTACACCAAGGCTACAGG 2954
Db 1539 GTCTACCTGCTCTCTACAACTCCGAGCCTGAGGAGGTTACACCAAGGCTACAGG 1598
QY 2955 CGACTACTACAACTTCCACTGCTGCGGAGCGCGCCCTCCGCGGCTCCCGCA 3014
Db 1599 CGACTACTACAACTTCCACTGCTGCGGAGCGCGCCCTCCGCGGCTCCCGCA 1658
QY 3015 TCCCCAGCTGCGATCAAGCTGGAGAACCGCTGGACTACGCGCAGCGCTGGGGGCTGCG 3074
Db 1659 TCCCCAGCTGCGATCAAGCTGGAGAACCGCTGGACTACGCGCAGCGCTGGGGGCTGCG 1718
QY 3075 GCGGGCGAGTGCCTATGGGACCTGGCGAGCCTGATGGCGGGGTGAGCGGGGAC 3134
Db 1719 GCGGGCGAGTGCCTATGGGACCTGGCGAGCCTGATGGCGGGGTGAGCGGGGAC 1778
QY 3135 CGGTTCTGGGTACCTCAGCGCGCTTCTCTATCTGGGCACACTCTCTTACAGCCGA 3194
Db 1779 CGGTTCTGGGTACCTCAGCGCGCTTCTCTATCTGGGCACACTCTCTTACAGCCGA 1838
QY 3195 AGAAGCCAGTTGTTATGAGCAGCTGTGTTGGTGGGGGCTGCT-----GCGGCGG 3245
Db 1839 AGAAGCCAGTTGTTATGAGCAGCTGTGTTGGTGGGGGCTGCT-----GCGGCGG 1898

QY 3246 CGSGCG 3305
Db 1899 CGSGCG 1958
QY 3306 AGCCCCCTACGGCTACACTCGGCCCGCTCAGGGGCTGCGGGCGCAGGAAAGCGACTTCAC 3365
Db 1959 AGCCCCCTACGGCTACACTCGGCCCGCTCAGGGGCTGCGGGCGCAGGAAAGCGACTTCAC 2018
QY 3366 CGCACTGATGTGTGGTACCTCGGCCCGCTGCGGAGAGTGGTGGAGAGTGGTGGAGTGGTGG 3425
Db 2019 CGCACTGATGTGTGGTACCTCGGCCCGCTGCGGAGAGTGGTGGAGAGTGGTGGAGTGGTGG 2078
QY 3426 TTGTGTCAAAAGCGAAATGGGCCCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3485
Db 2079 TTGTGTCAAAAGCGAAATGGGCCCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2138
QY 3486 GGGTTTGAGAGTGCAGGGGACCATGTTTGGCCATTGACTATTACTTTCCACCCAGAA 3545
Db 2139 GGGTTTGAGAGTGCAGGGGACCATGTTTGGCCATTGACTATTACTTTCCACCCAGAA 2198
QY 3546 GACCTGCTGATCTGTGGAGATGAAGCTTCTGGGTGTCACTATGAGCTTCACTATGTTGG 3605
Db 2199 GACCTGCTGATCTGTGGAGATGAAGCTTCTGGGTGTCACTATGAGCTTCACTATGTTGG 2258
QY 3606 AAGCTGCAAGTCTTCTTCAAAAGAGCGCTGAAGGAAACAGAAAGTACTGTGCGCCAG 3665
Db 2259 AAGCTGCAAGTCTTCTTCAAAAGAGCGCTGAAGGAAACAGAAAGTACTGTGCGCCAG 2318
QY 3666 CAGAAATGATTGCACTATTGATAAATTCGAAGGAAATTTGTCATCTGTGCTTGG 3725
Db 2319 CAGAAATGATTGCACTATTGATAAATTCGAAGGAAATTTGTCATCTGTGCTTGG 2378
QY 3726 GAAATGTTTGAAGCAGGATGACTCTGGAGCGCGGAGAGTGAAGAAATTTGTTAATCT 3785
Db 2379 GAAATGTTTGAAGCAGGATGACTCTGGAGCGCGGAGAGTGAAGAAATTTGTTAATCT 2438
QY 3786 GAAACTACAGGAGGAGGAGGCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3845
Db 2439 GAAACTACAGGAGGAGGAGGCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2498
QY 3846 GAAGCTGACAGTGTACACATTTGAAGGCTATGAATGTACAGCCATCTTTCGAAATSTCCT 3905
Db 2499 GAAGCTGACAGTGTACACATTTGAAGGCTATGAATGTACAGCCATCTTTCGAAATSTCCT 2558
QY 3906 GGAAGCCTTGAAGCAGGCTGAGTGTGCTGACACAGCAACACACAGCAGCAGCAGCAGCAG 3965
Db 2559 GGAAGCCTTGAAGCAGGCTGAGTGTGCTGACACAGCAACACACAGCAGCAGCAGCAGCAG 2618
QY 3966 TGACGCTTGTCTCTCTAGCCTCAATGAAGTGGAGAGAGAGCTTGTACAGTGGTCA 4025
Db 2619 TGACGCTTGTCTCTCTAGCCTCAATGAAGTGGAGAGAGAGCTTGTACAGTGGTCA 2678
QY 4026 GTGGGCAAGGCTTGCCTGCGCTTCCGAACTTACACGTGGAGCAGCAGTGTGAT 4085
Db 2679 GTGGGCAAGGCTTGCCTGCGCTTCCGAACTTACACGTGGAGCAGCAGTGTGAT 2738
QY 4086 TCAGTACTCCTGGATGGGCTCATGCTGTTGCCATGGCTGGCGATCCTTTCACCAATGT 4145
Db 2739 TCAGTACTCCTGGATGGGCTCATGCTGTTGCCATGGCTGGCGATCCTTTCACCAATGT 2798
QY 4146 CAACTCCAGGATGCTTACTTGGCCCTGATCTGTTTCAATGAGTACCGATGCA 4205
Db 2799 CAACTCCAGGATGCTTACTTGGCCCTGATCTGTTTCAATGAGTACCGATGCA 2858
QY 4206 GTCCCGGATGTACAGCAGTGTGCCGATGAGGAGCCTCTCTCAAGAGTGTGATGGCT 4265
Db 2859 GTCCCGGATGTACAGCAGTGTGCCGATGAGGAGCCTCTCTCAAGAGTGTGATGGCT 2918
QY 4266 CCAATACCCCCCAGGAAATTCCTGTGATGAAAGCACTGCTACTCTTCAAGCAATATTC 4325
Db 2919 CCAATACCCCCCAGGAAATTCCTGTGATGAAAGCACTGCTACTCTTCAAGCAATATTC 2978

QY 4326 AGTGGATGGCTGAAATAATCAAAATTTCTTGATGAATTCGAATGAACATACATCAAGGA 4385
Db 2979 AGTGGATGGCTGAAATAATCAAAATTTCTTGATGAATTCGAATGAACATACATCAAGGA 3038
QY 4386 ACTCGATCGTATCATGCGATGCAAAAGAAAATCCACATCTCCCTCAAGACGCTTCTA 4445
Db 3039 ACTCGATCGTATCATGCGATGCAAAAGAAAATCCACATCTCCCTCAAGACGCTTCTA 3098
QY 4446 CCAGCTCACCAGCTCCTCGATCCGCTGAGCCTATTCGAGAGAGCTGCATCAGTTTCA 4505
Db 3099 CCAGCTCACCAGCTCCTCGATCCGCTGAGCCTATTCGAGAGAGCTGCATCAGTTTCA 3158
QY 4506 TTTTGACCTGCTAATCAAGTACACATGATGAGCGTGACCTTCCGGAATGATGGCAGA 4565
Db 3159 TTTTGACCTGCTAATCAAGTACACATGATGAGCGTGACCTTCCGGAATGATGGCAGA 3218
QY 4566 GATCATCTCTGTCAGTGCCTGCAAGATCCTTCTGGGAAGTCAAGCCATCTATTTCCA 4625
Db 3219 GATCATCTCTGTCAGTGCCTGCAAGATCCTTCTGGGAAGTCAAGCCATCTATTTCCA 3278
QY 4626 CACCCAGTGAAGCATTTGGAACCCCTATTTCCGACCCAGCTCATGCCCCCTTTCAGATG 4685
Db 3279 CACCCAGTGAAGCATTTGGAACCCCTATTTCCGACCCAGCTCATGCCCCCTTTCAGATG 3338
QY 4686 TCTTCTGCTGTTATTAATCTGCACTACTCTCTGAGTGCCTTGGGAAATTTCTCTAT 4745
Db 3339 TCTTCTGCTGTTATTAATCTGCACTACTCTCTGAGTGCCTTGGGAAATTTCTCTAT 3398
QY 4746 TGATGTACAGTCTGTCATGAACATGTTCTCTGAATTTATTTGCTGGGCTTTTTTCTC 4805
Db 3399 TGATGTACAGTCTGTCATG-----GAATTTCTATTTGCTGGGCTTTTTTCTC 3447
QY 4806 TTTCTCTCCTTTCTTTTCT 4865
Db 3448 TTTCTCTCCTTTCTTTTCT 3507
QY 4866 TTTGCTTCCCATGTTGGCTCTATCTGTTTGAATGTTGTTGATGCTTTAAATCTG 4925
Db 3508 TTTGCTTCCCATGTTGGCTCTATCTGTTTGAATGTTGTTGATGCTTTAAATCTG 3567
QY 4926 TGATGATCCTCATATGCGCCAGTGTCAAGTGTGCTGTTTACAGCACTACTCTGTGCCA 4985
Db 3568 TGATGATCCTCATATGCGCCAGTGTCAAGTGTGCTGTTTACAGCACTACTCTGTGCCA 3627
QY 4986 GCCACACAAACGTTTACTTATCTTATGCGCGGAAGTTTAGAGAGCTAAGATATCTGG 5045
Db 3628 GCCACACAAACGTTTACTTATCTTATGCGCGGAAGTTTAGAGAGCTAAGATATCTGG 3687
QY 5046 GGAATCAAAACAAAA 5062
Db 3688 GGAATCAAAACAAAA 3704

RESULT 3
ID ABN95802
XX ABN95802 standard; DNA; 3715 BP.
XX AC
XX ABN95802;
XX 13-AUG-2002 (first entry)
XX DE
XX Gene #2300 used to diagnose liver cancer.
XX KW
XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumour; cytostatic; expression profile; disease state;
XX disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS
XX Homo sapiens.
XX PN
XX WO200229103-A2.
XX PD
XX 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30589.
XX 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX Diagnosing and detecting the progression of liver cancer;
XX hepatocellular carcinoma or metastatic liver tumor in a patient,
XX involves detecting the level of expression of two or more genes in a
XX liver tissue sample
XX Claim 1; SEQ ID NO 2300; 298pp; English.
XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumour in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 3715 BP; 841 A; 1055 C; 1001 G; 818 T; 0 other;

Query Match 70.8%; Score 3595.8; DB 24; Length 3715;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 3685; Conservative 0; Mismatches 7; Indels 45; Gaps 4;

QY 1335 TGCACGGGAGAGAACCTCTGTTTCCGCCACTCTCTCCACCTCTCTCTCTCTCTCTCTCTCC 1394
Db 4 TCCGGGGGAGAGAACCTCTGTTTCCGCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 63
QY 1395 CACCCGAGTGGGAGCCAGAGATCAAAAGATGAAAGGAGTCAAGTCTTCAGTAGCCA 1454
Db 64 CACCCGAGTGGGAG-CAGAGATCAAAAGATGAAAGGAGTCAAGTCTTCAGTAGCCA 122
QY 1455 AAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1514
Db 123 AAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 182
QY 1515 CTTATTTGCACCTACTTTCAGTGGACACTGAATTTGGAGGTGGAGGATTTGTTTTTTC 1574
Db 183 CTTATTTGCACCTACTTTCAGTGGACACTGAATTTGGAGGTGGAGGATTTGTTTTTTC 242
QY 1575 TTTTAAGATCTGGGATCTTTTGAATCTACCTTCAAGTATTAAAGACAGACTGTGAGC 1634
Db 243 TTTTAAGATCTGGGATCTTTTGAATCTACCTTCAAGTATTAAAGACAGACTGTGAGC 302
QY 1635 CTAGCAGGCGAGATCTTGTCCACCGTGTCTTCTTCTGACAGACTTTTGGAGCTGTCA 1694
Db 303 CTAGCAGGCGAGATCTTGTCCACCGTGTCTTCTTCTGACAGACTTTTGGAGCTGTCA 362
QY 1695 GAGCGCTTTTTCGCTGTTGCTCCCGCAAGTTTCTCTCTGAGCTTCCCGCAGTGGG 1754
Db 363 GAGCGCTTTTTCGCTGTTGCTCCCGCAAGTTTCTCTCTGAGCTTCCCGCAGTGGG 422
QY 1755 CAGCTAGCTGAGCGACTACCGCATCATCAGGCTGTGGAACCTCTTCTGAGCAAGAGAA 1814
Db 423 CAGCTAGCTGAGCGACTACCGCATCATCAGGCTGTGGAACCTCTTCTGAGCAAGAGAA 482
QY 1815 GGGGAGCGGGTAAAGGGAAGTAGGTGGAAGATTACGCCAAGCTCAAGGATGGAGTGCA 1874

Db 483 GGGAGGCGGGTAAAGGAAGTAGTGGGAAGATTACCAAGCTCAAGGATGAAGTGA 542
QY 1875 GTTAGGGCTGGGAAGGCTTACCTCGCGCGCGTCCAAAGACCTACCGAGGAGCTTTCCA 1934
Db 543 GTTAGGGCTGGGAAGGCTTACCTCGCGCGCGTCCAAAGACCTACCGAGGAGCTTTCCA 602
QY 1935 GAATCTCTCCAGAGCGTGGCCGAAGTAGATCCAGAACCCGGGGCCCGAGAGGC 1994
Db 603 GAATCTCTCCAGAGCGTGGCCGAAGTAGATCCAGAACCCGGGGCCCGAGAGGC 662
QY 1995 CGGAGCGCAGCACCTCCCGCGCCCAAGTTTGTCTGTCTGTCAGCAGCAGCAGCAGCA 2054
Db 663 CGGAGCGCAGCACCTCCCGCGCCCAAGTTTGTCTGTCT----- 701
QY 2055 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2114
Db 702 ---GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 758
QY 2115 TAGCCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2174
Db 759 TAGCCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 818
QY 2175 AGGCCCCACAGGCTACCTGGTCTCTGGATGAGGAACAGCAACCTTCACAGCCGCTGCGC 2234
Db 819 AGGCCCCACAGGCTACCTGGTCTCTGGATGAGGAACAGCAACCTTCACAGCCGCTGCGC 878
QY 2235 CCTGGAGTCCACCCCGAGAGGTTGGTCCAGAGCCTGAGAGCCGCGCTGGCGCCAG 2294
Db 879 CCTGGAGTCCACCCCGAGAGGTTGGTCCAGAGCCTGAGAGCCGCGCTGGCGCCAG 938
QY 2295 CAAGGGGCTGCCGACAGCTGCCAGCACCTCCCGACAGGATGACTCAGCTGCCCATC 2354
Db 939 CAAGGGGCTGCCGACAGCTGCCAGCACCTCCCGACAGGATGACTCAGCTGCCCATC 998
QY 2355 CAGTTGTCCCTGCTGGGCCCCACCTTCCCGGGCTTAAGCAGCTGCTCCGCTGACCTTAA 2414
Db 999 CAGTTGTCCCTGCTGGGCCCCACCTTCCCGGGCTTAAGCAGCTGCTCCGCTGACCTTAA 1058
QY 2415 AGACATCTTAGCGAGGCGCAGCACCATCTCTTCCAGCAAGCAGCAGCAGCAGCAGT 2474
Db 1059 AGACATCTTAGCGAGGCGCAGCACCATCTCTTCCAGCAAGCAGCAGCAGCAGCAGT 1118
QY 2475 ATCCGAGGCGCAGCAGCGGAGAGCGAGGAGGCGCTCGGGGCTCCCATCTCCCAA 2534
Db 1119 ATCCGAGGCGCAGCAGCGGAGAGCGAGGAGGCGCTCGGGGCTCCCATCTCCCAA 1178
QY 2535 GGACAATTACTTAGGGGCACTTCGACCATTTCTGACAAAGCCAGAGGTTGTGAAGC 2594
Db 1179 GGACAATTACTTAGGGGCACTTCGACCATTTCTGACAAAGCCAGAGGTTGTGAAGC 1238
QY 2595 AGTGTGGTGTCCATGGGCCCTGGGTGTGGAGCGCTTGGAGCATCTGAGTCCAGGGGAACA 2654
Db 1239 AGTGTGGTGTCCATGGGCCCTGGGTGTGGAGCGCTTGGAGCATCTGAGTCCAGGGGAACA 1298
QY 2655 GCTTCGGGGGATGTCATGACGCCCACTTTGGGAGTTCCACCGCTGTGCGTCCCAAC 2714
Db 1299 GCTTCGGGGGATGTCATGACGCCCACTTTGGGAGTTCCACCGCTGTGCGTCCCAAC 1358
QY 2715 TCTTTGTCCCAATTGGCCGAATGCAAGGTTCTCTCTAGACAGCAGCAGCAGCAGCAG 2774
Db 1359 TCTTTGTCCCAATTGGCCGAATGCAAGGTTCTCTCTAGACAGCAGCAGCAGCAGCAG 1418
QY 2775 CACTGAAGATGCTGCTGAGTATTCCTTTCAAGGGAGGTTACACAAAGGCTAGAAGG 2834
Db 1419 CACTGAAGATGCTGCTGAGTATTCCTTTCAAGGGAGGTTACACAAAGGCTAGAAGG 1478
QY 2835 CGAGAGCCTAGGCTGCTGCGAGCGCTGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2894
Db 1479 CGAGAGCCTAGGCTGCTGCGAGCGCTGCGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1538
QY 2895 GTCTACCCCTGTCTCTACAGTCCGAGCAGCTGGAGGAGGAGCTGCGCTACAGAGTCG 2954
Db 1539 GTCTACCCCTGTCTCTACAGTCCGAGCAGCTGGAGGAGGAGCTGCGCTACAGAGTCG 1598

QY 2955 CGACTACTACAACCTTTCCACTGGCTGTGGCCGGAACCGCGCCCTCGCGCGCTCCCA 3014
Db 1599 CGACTACTACAACCTTTCCACTGGCTGTGGCCGGAACCGCGCCCTCGCGCGCTCCCA 1658
QY 3015 TCCCCAGCGCTCGCATCAAGCTGAGAACCCGCTGGACTACGACAGCGCTAGGCCTAGGC 3074
Db 1659 TCCCCAGCGCTCGCATCAAGCTGAGAACCCGCTGGACTACGACAGCGCTAGGCCTAGGC 1718
QY 3075 GCGCGCGCAGTGGCGCTATGGGACCTTGGCAGCCTGTGATGGCGGGGTGAGCGGAC 3134
Db 1719 GCGCGCGCAGTGGCGCTATGGGACCTTGGCAGCCTGTGATGGCGGGGTGAGCGGAC 1778
QY 3135 CGGTTCTGGGTCACCTTCAGCCGCGCTTCCTCATCTCGGCACACTCTCTTTCACAGCGA 3194
Db 1779 CGGTTCTGGGTCACCTTCAGCCGCGCTTCCTCATCTCGGCACACTCTCTTTCACAGCGA 1838
QY 3195 AGAAGCCCAAGTTGTATGGACCGCTGTGGTGGGGGTGTGGTGGCGCGCGCGG 3245
Db 1839 AGAAGCCCAAGTTGTATGGACCGCTGTGGTGGGGGTGTGGTGGCGCGCGCGG 1898
QY 3245 CGG 3305
Db 1899 CGG 1958
QY 3306 AGCCCCCTACGGCTACACTCGGCCCTCAGGGGTGGCGGCGCGCGCGCGCGCGCGG 3365
Db 1959 AGCCCCCTACGGCTACACTCGGCCCTCAGGGGTGGCGGCGCGCGCGCGCGCGG 2018
QY 3366 CGCAGCTGATGTGGTACCTCGCGGCGATGGTGAAGAGTGGCGCTATCCAGTCCAC 3425
Db 2019 CGCAGCTGATGTGGTACCTCGCGGCGATGGTGAAGAGTGGCGCTATCCAGTCCAC 2078
QY 3426 TTGTGTCAAAAGCGAAATGGGCCCTGGATGGATAGCTTCTCCGGACCTTACGGGGACAT 3485
Db 2079 TTGTGTCAAAAGCGAAATGGGCCCTGGATGGATAGCTTCTCCGGACCTTACGGGGACAT 2138
QY 3486 GCGTTTGGAGACTGCGAGGACCATGTTTGCCTATTTGCTTATTTACCTTCCACCCAGAA 3545
Db 2139 GCGTTTGGAGACTGCGAGGACCATGTTTGCCTATTTGCTTATTTACCTTCCACCCAGAA 2198
QY 3546 GACCTGCTGATCTGTGGAGATGAAGCTCTGGGTGTCACTATGGAGCTCTACATGTGG 3605
Db 2199 GACCTGCTGATCTGTGGAGATGAAGCTCTGGGTGTCACTATGGAGCTCTACATGTGG 2258
QY 3606 AAGCTGCAAGTCTTCTTCAAAAGAGCGCTGAAGGGAACACAGAGTACCTGTCGCCAG 3665
Db 2259 AAGCTGCAAGTCTTCTTCAAAAGAGCGCTGAAGGGAACACAGAGTACCTGTCGCCAG 2318
QY 3666 CAGAAATGATGCACTATTGATAAATTCGAAAGGAAAAATTTGCTTGTCTCTCG 3725
Db 2319 CAGAAATGATGCACTATTGATAAATTCGAAAGGAAAAATTTGCTTGTCTCTCG 2378
QY 3726 GAAATGTTATGAAGCAGGATGACTCTGGAGCGCGGAGCTCAAGAAACTTGGTAACT 3785
Db 2379 GAAATGTTATGAAGCAGGATGACTCTGGAGCGCGGAGCTCAAGAAACTTGGTAACT 2438
QY 3786 GAACTACAGGAGGAGGAGGCTTCCAGCACCGAGGCTGAGGAGGAGGAGGAGGAGG 3845
Db 2439 GAACTACAGGAGGAGGAGGCTTCCAGCACCGAGGCTGAGGAGGAGGAGGAGGAGG 2498
QY 3846 GAAGCTACAGGCTGTCACACATTTGAAGCTATGAATGTAGCCCATCTTTCTGAATGTCT 3905
Db 2499 GAAGCTACAGGCTGTCACACATTTGAAGCTATGAATGTAGCCCATCTTTCTGAATGTCT 2558
QY 3906 GGAAGCATTGAGCAGGCTAGTGTGTGGACAGCAGCAACACCGAGCTCCCT 3965
Db 2559 GGAAGCATTGAGCAGGCTAGTGTGTGGACAGCAGCAACACCGAGCTCCCT 2618
QY 3966 TGAGCCTTCTCTAGGCTCAATGAAGTGGGAGAGAGAGCTTGTACAGTGGTCAA 4025
Db 2619 TGAGCCTTCTCTAGGCTCAATGAAGTGGGAGAGAGAGCTTGTACAGTGGTCAA 2678

QY	4026	GTGGGCAAGGCCCTTGCTGGCTTCCGCAAACTTACAGCTGGACGACAGATGGCTGTCAT	4085
DB	2679	GTGGGCCAAGGCCCTTGCTTCCGCAAACTTACAGCTGGACGACAGATGGCTGTCAT	2738
QY	4086	TCAGTACTCCTGGATGGGGCTCATGTTGGTGGCATGGCTGGCATCCTTCACCAATGT	4145
DB	2739	TCAGTACTCCTGGATGGGGCTCATGTTGGTGGCATGGCTGGCGATCTCTCCACCAATGT	2798
QY	4146	CAACTCCAGATGCTCTACTTTCGCCCTCGATCTGGTTTCAATAGTACCGCATCACAA	4205
DB	2799	CAACTCCAGATGCTCTACTTTCGCCCTCGATCTGGTTTCAATAGTACCGCATCACAA	2858
QY	4206	GTCCCGGATGTACGCCAGTGTGTCGGATGAGGCCACCTCTCTCAAGATTTGGATGGCT	4265
DB	2859	GTCCCGGATGTACGCCAGTGTGTCGGATGAGGCCACCTCTCTCAAGATTTGGATGGCT	2918
QY	4266	CCAAATCACCCCCAGGAATTCCTGTGCATGAAGACACTGCTACTCTTCAGCATTTATCC	4325
DB	2919	CCAAATCACCCCCAGGAATTCCTGTGCATGAAGACACTGCTACTCTTCAGCATTTATCC	2978
QY	4326	AGTGGATGGGCTGAAAAATCAAAAATCTTTTGATGAACCTTCGAATGAACATCAAGGA	4385
DB	2979	AGTGGATGGGCTGAAAAATCAAAAATCTTTTGATGAACCTTCGAATGAACATCAAGGA	3038
QY	4386	ACTCGATCTGATCATTTGATGATCAAAAGAAAAAATCCACACATCTCTGCTCAAGACGCTTCTA	4445
DB	3039	ACTCGATCTGATCATTTGATGATCAAAAGAAAAAATCCACACATCTCTGCTCAAGACGCTTCTA	3098
QY	4446	CCAGCTCACCAAGCTCCTGGATCCGCTGAGCGCTATTTCGGAGAGAGCTGCATCAGTTTCA	4505
DB	3099	CCAGCTCACCAAGCTCCTGGATCCGCTGAGCGCTATTTCGGAGAGAGCTGCATCAGTTTCA	3158
QY	4506	TTTTGACCTGTAATCAAGTCAACATGGTGAGGCTGGACTTTCCGGAATGATGCGAGA	4565
DB	3159	TTTTGACCTGTAATCAAGTCAACATGGTGAGGCTGGACTTTCCGGAATGATGCGAGA	3218
QY	4566	GATCATCTCTGTGCAAGTCCCAAGATCCTTCTTGGGAAAGTCAAGCCCATCTATTTCACA	4625
DB	3219	GATCATCTCTGTGCAAGTCCCAAGATCCTTCTTGGGAAAGTCAAGCCCATCTATTTCACA	3278
QY	4626	CACCCAGTGAAGCATTTGGAACCCCTATTTCGCCACCCAGCTCATGCCCTTTTCAGATG	4685
DB	3279	CACCCAGTGAAGCATTTGGAACCCCTATTTCGCCACCCAGCTCATGCCCTTTTCAGATG	3338
QY	4686	TCCTCTGCTGTTAATACCTGCAGCTACTCTCTGCAGTGCCCTGGGGAATTTCTCTCTAT	4745
DB	3339	TCCTCTGCTGTTAATACCTGCAGCTACTCTCTGCAGTGCCCTGGGGAATTTCTCTCTAT	3398
QY	4746	TGATGTACAGTCTCATGAACATGTTCTGAAATCTATTTCGTGGGCTTTTTTTTCTC	4805
DB	3399	TGATGTACAGTCTCATGAAGTCTGTTTGGTGTGTTGATGCGCTTTAAATCTG	3447
QY	4806	TTTCTCTCCTTTCTTTTCTTCTTCCCTCCCTATCTAAACCTCCCATGGCACCTTCAGAC	4865
DB	3448	TTTCTCTCCTTTCTTTTCTTCTTCCCTCCCTATCTAAACCTCCCATGGCACCTTCAGAC	3507
QY	4866	TTTGTCTCCCATTTGTCGCTCCTATCTGTGTTTGAATGGTGTGTTATGCCCTTTAAATCTG	4925
DB	3508	TTTGTCTCCCATTTGTCGCTCCTATCTGTGTTTGAATGGTGTGTTATGCCCTTTAAATCTG	3567
QY	4926	TCATCATCCTCATATGCCAGTGTCAAGTTGTGCTTGTGTTTACAGCAGTACTCTGTGCCA	4985
DB	3568	TCATCATCCTCATATGCCAGTGTCAAGTTGTGCTTGTGTTTACAGCAGTACTCTGTGCCA	3627
QY	4986	GCCACACAAAGTTTACTTATCTTATGTCACGGGAAGTTTAGAGAGCTTAAGATTATCTGG	5045
DB	3628	GCCACACAAAGTTTACTTATCTTATGTCACGGGAAGTTTAGAGAGCTTAAGATTATCTGG	3687
QY	5046	GGAAATCAAAACAAAAA	5062
DB	3688	GGAAATCAAAACAAAAA	3704

RESULT 4	
AAQ12001	AAQ12001 standard; cDNA; 3715 BP.
XX AC	
XX AC	AAQ12001;
XX DT	20-AUG-1991 (first entry)
XX XX	Full-length human androgen receptor coding sequence.
DE DE	hAR; DNA-binding protein; steroid hormone; ss.
XX KW	Homo sapiens.
XX OS	
XX FH	Key Location/Qualifiers
CDS	532..3288
FT FT	/tag= a
FT FT	/product= full-length 918 residue hAR
FT FT	/note= "includes shorter 734 residue hAR"
XX XX	
PN PN	W09107423-A.
XX PD	30-MAY-1991.
XX XX	
PX PF	19-OCT-1990; 90WO-US06015.
XX PR	17-NOV-1989; 89US-0438775.
XX PA	(ARCH-) ARCH DEV CORP.
XX PI	Liao S, Chang C;
XX XX	
DR DR	WPI; 1991-178048/24.
DR P-	PSDB; AAR12223.
XX PT	Androgen receptor and RX2 DNA binding proteins - DNA sequ
PT PT	and antibodies for detection and quantification methods
XX PS	Claim 4; Fig 3; 79pp; English.
XX CC	This sequence was isolated by screening commercially avail
CC CC	testis and prostate lambda gtlI cDNA libraries. Initial s
CC CC	was with probes designed for homology to nucleotide sequen
CC CC	DNA-binding domain of known steroid receptors. Positive c
CC CC	then screened with 24mer probes specific for the various
CC CC	receptors to eliminate those which coded for known recep
CC CC	remaining clones were analysed by restriction mapping and
CC CC	sequenced. The human AR coding sequence is given here.
XX SX	Sequence 3715 BP; 842 A; 1055 C; 1003 G; 815 T; 0 other;
Query Match	70.7%; Score 3594.2; DB 12; Leng
Best Local Similarity	98.6%; Pred. No. 0;
Matches 3684;	Conservative 0; Mismatches 8; Indels
OY 1335	TGCACGGCGGAGAACCCCTGTGTTTCCCCCCTCTCTCTCCACCCTCCCTCC
Db 4	TTCGGCGGAGAACCCCTGTGTTTCCCCCCTCTCTCTCTCCACCCTCCCTCC
OY 1395	CACCCCGAGTGCGGAGCCAGAGATCAAAAGATGAAAAGGCAGTCAGTCTT
Db 64	CACCCCGAGTGCGGAG-CAGAGATCAAAAGATGAAAAGGCAGTCAGTCTT
OY 1455	AAAAACAAAACAACAAAAAACAAGCCGAATAAAGAAGAAAAGATAT
Db 123	AAAAACAAAACAACAAAAAACAAGCCGAATAAAGAAGAAAAGATAT
OY 1515	CTTATTTCGACCTACTTCAGTGGACACTGAATTTGGAAGTGGCAGGATTTT
Db 183	CTTATTTCGACCTACTTCAGTGGACACTGAATTTGGAAGTGGCAGGATTTT
OY 1575	TTTTTAAGATCTGGGCATCTTTTGAATCTACCCCTCAAGTATTTAGACACAG

Db	3508	TTTTGCTTCCCATTTGGCTCCTATCTGTGTTTGAATGGTCTGTATGCCCTTTAAATCTG	3567
Qy	4926	TGATGATCCTCATATGCGCCAGTCAGTGTGCTTGTTTACAGCACTACTCTGTGCCA	4985
Db	3568	TGATGATCCTCATATGCGCCAGTCAGTGTGCTTGTTTACAGCACTACTCTGTGCCA	3627
Qy	4986	GCCACACAAACGTTTACTTATCTTATGCCACGGGAAGTTTAGAGAGCTTAAGATTATCTGG	5045
Db	3628	GCCACACAAACGTTTACTTATCTTATGCCACGGGAAGTTTAGAGAGCTTAAGATTATCTGG	3687
Qy	5046	GGAAATCAAAACAAAAA 5062	
Db	3688	GGAAATCAAAACAAAAA 3704	
RESULT 5			
AAAT63407	ID	AAAT63407 standard; cDNA; 3569 BP.	
XX	XX	AAAT63407;	
XX	XX	22-JUN-1997 (first entry)	
XX	XX	Androgen receptor cDNA.	
XX	XX	Androgen receptor; acidic fibroblast growth factor; aFGF;	
KW	KW	antisense; Benign prostatic hyperplasia; prostate cancer; therapy;	
KW	KW	ds.	
XX	OS	Homo sapiens.	
XX	XX	Location/Qualifiers	
PH	Key	363..3122	
FT	CDS	/*tag= a	
FT	FT	complement (916..936)	
FT	FT	/*tag= b	
FT	FT	/note= "antisense oligonucleotide preferred for	
FT	FT	use in methods of the invention"	
FT	FT	complement (927..947)	
FT	FT	/*tag= c	
FT	FT	/note= "antisense oligonucleotide preferred for	
FT	FT	use in methods of the invention"	
FT	FT	complement (927..936)	
FT	FT	/*tag= d	
FT	FT	/note= "antisense oligonucleotide preferred for	
FT	FT	use in methods of the invention"	
XX	XX	WO9711170-A1.	
XX	XX	27-MAR-1997.	
XX	XX	20-SEP-1996; 96WO-US15081.	
XX	XX	20-SEP-1995; 95US-0004018.	
XX	XX	(WORC-) WORCESTER FOUND BIOMEDICAL RES.	
PA	PA	Zamecnik PA;	
PI	PI	WPI; 1997-202879/18.	
XX	XX	P-PSDB; AAW14783.	
DR	DR	Oligonucleotide(s) antisense to human androgen receptor and acidic	
XX	XX	FGF genes - used to inhibit gene expression, for the treatment of	
PT	PT	benign prostatic hyperplasia	
PT	PT	Claim 2; Page 21-29; 51pp; English.	
XX	XX	A cDNA clone (AAAT63407) codes for the human androgen receptor	
CC	CC	(AAW14783). Methods of selectively inhibiting the growth, or of	
CC	CC	killing, prostatic cells involve the use of antisense	
CC	CC	oligonucleotides (see also AAT63200, AAT63404-05) to this androgen	

QY	4622	TCACACCCAGTGAACCACTTGGAAACCCCTATTATTTCCCAACCCAGCTCATGCCCCCTTTCA	4681
Db	3109	TCACACCCAGTGAACCACTTGGAAACCCCTATTATTTCCCAACCCAGCTCATGCCCCCTTTCA	3168
QY	4682	GATGCTCTTCGCCCTGTTATAACTCTGCAGTACTCTCTCTGCAGTGCCTTGGGGAATTCCT	4741
Db	3169	GATGCTCTTCGCCCTGTTATAACTCTGCAGTACTCTCTCTGCAGTGCCTTGGGGAATTCCT	3228
QY	4742	CTATTGATGTPACAGCTGTGTCATGAACATGTTCCCTGAAATTCCTATTGTCTGGGCTTTTTTT	4801
Db	3229	CTATTGATGTPACAGCTGTGTCATGAACATGTTCCCTGAAATTCCTATTGTCTGGGCTTTTTTT	3288
QY	4802	TCTCTTTCTCTCCCTTTCTTTTTTCTTCCCTCCCTATCTAACCCCTCCCATGGCACCTTC	4861
Db	3289	TCTCTTTCTCTCCCTTTCTTTTTTCTTCCCTCCCTATCTAACCCCTCCCATGGCACCTTC	3348
QY	4862	AGACTTTGCTTCCCATGTGGCTCCCTATCTGTGTTTTGAATGGTGTGTTATGCGCTTTTAAA	4921
Db	3349	AGACTTTGCTTCCCATGTGGCTCCCTATCTGTGTTTTGAATGGTGTGTTATGCGCTTTTAAA	3408
QY	4922	TCTGTGATGATCCCTCATATATGCCCCAGTGTCAAGTTGTGCTTTTACAGCACTACTCTGT	4981
Db	3409	TCTGTGATGATCCCTCATATATGCCCCAGTGTCAAGTTGTGCTTTTACAGCACTACTCTGT	3468
QY	4982	GCCAGCCACACAAACGTTTACTTATCTTATGCCACGGGAAGTTTACAGAGCTTAAGATTAT	5041
Db	3469	GCCAGCCACACAAACGTTTACTTATCTTATGCCACGGGAAGTTTACAGAGCTTAAGATTAT	3528
QY	5042	CTGGGGAATCAAAACAAAAAACAAGCAACAAAAA	5082
Db	3529	CTGGGGAATCAAAACAAAAAACAAGCAACAAAAA	3569
RESULT 6			
AA91772			
ID	AA91772 standard; cDNA; 3569 BP.		
XX	AA91772;		
XX	19-MAR-1990 (first entry)		
XX	Human androgen receptor cDNA.		
XX	Human androgen receptor; monoclonal antibody; ployclonal antibody;		
XX	cancer; probe.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	CDS	363..3122	
FT		/*tag= a	
XX	W08909791-A.		
XX	19-OCT-1989.		
XX	13-APR-1989;	89WO-US01548.	
XX	14-APR-1988;	88US-0182646.	
XX	(UYN-) UNIVERSITY OF NORTH CAROLINA.		
XX	French FS, Wilson EM, Joseph DR, Lubahn DB;		
XX	WPI: 1989-324206/44.		
XX	P-PSDB; AAP93109.		
XX	DNA encoding androgen receptor protein - useful for transforming		
XX	eukaryotic hosts for protein expression and subsequent antibody prodn		
XX	Disclosure; Fig. 4; 41pp; English.		
XX	Complementary DNA sequences derived from the cDNA may be used as prob		

CC detect the presence of androgen receptor (AR mRNA in tumour cells, and to
 CC detect AR gene defects using DNA hybridisation assays.
 XX

Sequence 3569 BP; 796 A; 1008 C; 975 G; 790 T; 0 other;

Query Match 69.8%; Score 3545.4; DB 10; Length 3569;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 3568; Conservative 0; Mismatches 1; Indels 12; Gaps 1;

QY 1502 TAATAACTCAGTCTTATTTGACCTACTTACAGTGACACTGAATTTGAAGTGGAGGA 1561
 DB 1 TAATAACTCAGTCTTATTTGACCTACTTACAGTGACACTGAATTTGAAGTGGAGGA 60
 QY 1562 TTTTGTGTTTTTCTTTTAAAGATCTGGCACTCTTTTGAATCTACCTTCAAGTATTAAAG 1621
 DB 61 TTTTGTGTTTTTCTTTTAAAGATCTGGCACTCTTTTGAATCTACCTTCAAGTATTAAAG 120
 QY 1622 ACAGACTGTGAGCTAGCAGGAGAGATCTGTCCACCGTGTCTTCTTCGACAGAGAC 1681
 DB 121 ACAGACTGTGAGCTAGCAGGAGAGATCTGTCCACCGTGTCTTCTTCGACAGAGAC 180
 QY 1682 TTTGAGGCTCTCAGAGCGCTTTTGGTGGTGTCTCCCGCAAGTTTCCCTTCTCTGGAGCT 1741
 DB 181 TTTGAGGCTCTCAGAGCGCTTTTGGTGGTGTCTCCCGCAAGTTTCCCTTCTCTGGAGCT 240
 QY 1742 TCCCGCAGGTGGCAGCTAGCTGCAGGACTACCGCATCATCAGCCCTGTTGAACCTTT 1801
 DB 241 TCCCGCAGGTGGCAGCTAGCTGCAGGACTACCGCATCATCAGCCCTGTTGAACCTTT 300
 QY 1802 CTGAGCAAGAGAGGGAGCGGGTAAAGGAGTAAAGTGAAGTTCAGGCAAGCTCAA 1861
 DB 301 CTGAGCAAGAGAGGGAGCGGGTAAAGGAGTAAAGTGAAGTTCAGGCAAGCTCAA 360
 QY 1862 GGATGAAGTGCAGTTAGGGCTGGGAGGTTCTACCTCGCGCGCGTCCAAAGACTACC 1921
 DB 361 GGATGAAGTGCAGTTAGGGCTGGGAGGTTCTACCTCGCGCGCGTCCAAAGACTACC 420
 QY 1922 GAGGAGCTTTCCAGAACTCTGTCAGAGCTGCGGAGTGTATCCAGAACCCGGGCCCA 1981
 DB 421 GAGGAGCTTTCCAGAACTCTGTCAGAGCTGCGGAGTGTATCCAGAACCCGGGCCCA 480
 QY 1982 GGCACCCAGAGCGCGAGCGAGCACTTCCCGCGCGCGAGTTTGTGTGTGTGTCAGAGC 2041
 DB 481 GGCACCCAGAGCGCGAGCGAGCACTTCCCGCGCGCGAGTTTGTGTGTGTGTCAGAGC 540
 QY 2042 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2101
 DB 541 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 588
 QY 2102 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2161
 DB 589 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 648
 QY 2162 AAGCCATCGTAGAGGCCCCAGAGCTACCTGGTCTTGGATGAGGAACAGCACTTCCAC 2221
 DB 649 AAGCCATCGTAGAGGCCCCAGAGCTACCTGGTCTTGGATGAGGAACAGCACTTCCAC 708
 QY 2222 AGCCGAGTCGSCCTGAGTCCACCCGAGAGAGTGGTCCAGAGCCTGGAGCGG 2281
 DB 709 AGCCGAGTCGSCCTGAGTCCACCCGAGAGAGTGGTCCAGAGCCTGGAGCGG 768
 QY 2282 CCGTGGCCGCCAGCAAGGGGTGCGCAGCAGCTGCCAGCACTTCCGAGCAGCATGACT 2341
 DB 769 CCGTGGCCGCCAGCAAGGGGTGCGCAGCAGCTGCCAGCACTTCCGAGCAGCATGACT 828
 QY 2342 CAGCTGCCCCATCCAGTGTGCTGTGGGCCCACTTTTCCCGGGCTTAAAGCAGCTGCT 2401
 DB 829 CAGCTGCCCCATCCAGTGTGCTGTGGGCCCACTTTTCCCGGGCTTAAAGCAGCTGCT 888
 QY 2402 CCGCTGACCTTAAAGACATCTGAGCGAGGCGCAGCAGCATGCACTCTTCAGCAACAGC 2461
 DB 889 CCGCTGACCTTAAAGACATCTGAGCGAGGCGCAGCAGCATGCACTCTTCAGCAACAGC 948

QY 2462 AGCAGGAAGCAGTATCCGAAGGAGCAGCAGCAGCGGAGAGCGAGGAGCCTCGGGGGCTC 2521
 DB 949 AGCAGGAAGCAGTATCCGAAGGAGCAGCAGCAGCGGAGAGCGAGGAGCCTCGGGGGCTC 1008
 QY 2522 CCACCTTCTCCAAAGGAGCAATTTAGGGGGCAGCTTCAGCAATTTCTGACAGCGCAAGG 2581
 DB 1009 CCACCTTCTCCAAAGGAGCAATTTAGGGGGCAGCTTCAGCAATTTCTGACAGCGCAAGG 1068
 QY 2582 AGTTGTGAAGCAGTGTGGTGTCCATGGGCTGGTGTGGAGCGTTGGAGCATCTGA 2641
 DB 1069 AGTTGTGAAGCAGTGTGGTGTCCATGGGCTGGTGTGGAGCGTTGGAGCATCTGA 1128
 QY 2642 GTCCAGGGGAGCAGCTTTCGGGGGGAATTCATGTACGCCCACTTTTGGGAGTTCCACCG 2701
 DB 1129 GTCCAGGGGAGCAGCTTTCGGGGGGAATTCATGTACGCCCACTTTTGGGAGTTCCACCG 1188
 QY 2702 CTGTGCTGCCACTCTTGTGCCCCCATTTGGCCGAATGCAAGGTTCTCTGTACAGGACA 2761
 DB 1189 CTGTGCTGCCACTCTTGTGCCCCCATTTGGCCGAATGCAAGGTTCTCTGTACAGGACA 1248
 QY 2762 GCGCAGGCAAGAGCACTGCTGAGTATTCCTCTTCAAGGAGGTTTACACCA 2821
 DB 1249 GCGCAGGCAAGAGCACTGCTGAGTATTCCTCTTCAAGGAGGTTTACACCA 1308
 QY 2822 AAGGCTAGAAGGCGAGAGCCTAGGCTGCTGTGCGAGCGTGCAGAGGAGCTCCGGGA 2881
 DB 1309 AAGGCTAGAAGGCGAGAGCCTAGGCTGCTGTGCGAGCGTGCAGAGGAGCTCCGGGA 1368
 QY 2882 CACTTGAAGTCCCTTACCCTGTCTCTTACAAGTCCGGAGCACTGGACAGGAGCTG 2941
 DB 1369 CACTTGAAGTCCCTTACCCTGTCTCTTACAAGTCCGGAGCACTGGACAGGAGCTG 1428
 QY 2942 CGTACAGAGTCGAGCACTACTACAACTTTCGACTGGCTGTGGCGGAGCCGCCCTC 3001
 DB 1429 CGTACAGAGTCGAGCACTACTACAACTTTCGACTGGCTGTGGCGGAGCCGCCCTC 1488
 QY 3002 CGCGCCTCCCATCCCGAGCCTCGCATCAAGCTGGAGAACCCGCTGGACTAGGCGAGCG 3061
 DB 1489 CGCGCCTCCCATCCCGAGCCTCGCATCAAGCTGGAGAACCCGCTGGACTAGGCGAGCG 1548
 QY 3062 CTTGGGCGGTGCGGCGGCGAGTGGCGCTATGGGAGCTTGGGAGCTTCATGGCGCGG 3121
 DB 1549 CTTGGGCGGTGCGGCGGCGAGTGGCGCTATGGGAGCTTGGGAGCTTCATGGCGCGG 1608
 QY 3122 GTGAGCGGAGCGCGGTTCCTGGGTCACCTCAGCCGCGCTTCTCATCTTGGCAGACTC 3181
 DB 1609 GTGAGCGGAGCGCGGTTCCTGGGTCACCTCAGCCCTCAGCGGCTTGGCGGCGGAG 1668
 QY 3182 TCTTCAGCGCGAGAGGCGAGTGTGTATGACCGCTGTGTGTGTGTGGGGTGGTGGCG 3241
 DB 1669 TCTTCAGCGCGAGAGGCGAGTGTGTATGACCGCTGTGTGTGTGTGGGGTGGTGGCG 1728
 QY 3242 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 3301
 DB 1729 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 1788
 QY 3302 CTGTAGCCCCCTACGGCTACACTCGGCCCTCAGGGGCTTGGCGGCGCAGGAAGGAGCT 3361
 DB 1789 CTGTAGCCCCCTACGGCTACACTCGGCCCTCAGGGGCTTGGCGGCGCAGGAAGGAGCT 1848
 QY 3362 TCACCGCAGCTGATGTGTGTACCTTGGGCGCATGTGTAGCAGAGTGCCTATCCAGTGC 3421
 DB 1849 TCACCGCAGCTGATGT 1908
 QY 3422 CCACCTTGTGTCAAAAGGAGAAATGGGCCCTTGGATGATAGTACTCCGACCTTACGGGG 3481
 DB 1909 CCACCTTGTGTCAAAAGGAGAAATGGGCCCTTGGATGATAGTACTCCGACCTTACGGGG 1968
 QY 3482 ACATGCTTTGGAGACTGCGCAGGAGCACTTTTGGCCCATTTGACTATTACTTTCCACCCC 3541
 DB 1969 ACATGCTTTGGAGACTGCGCAGGAGCACTTTTGGCCCATTTGACTATTACTTTCCACCCC 2028
 QY 3542 AGAAGACCTGCCTGTGTGTGGAGATGAAGCTTCTGGGTGTGCACTATGAGAGCTCTACAT 3601

Db 2029 AGAAGACCTGCTGATCTCTGGAGATGAAGCTCTGGGTGTCACATATGGAGCTCTCACAT 2088
 QY 3602 GTGGAAGCTGCAAGGCTCTCTTCAAAAGAGCCGCTGAAGGGAACAGAAAGTACCTGTGCG 3661
 Db 2089 GTGGAAGCTGCAAGGCTCTCTTCAAAAGAGCCGCTGAAGGGAACAGAAAGTACCTGTGCG 2148
 QY 3662 CGAGCAAGATGATTCACATATTTGATAATTCGGAAGGAAATTTGCCATCTTCTGCTG 3721
 Db 2149 CCAGCAAGATGATTCACATATTTGATAATTCGGAAGGAAATTTGCCATCTTCTGCTG 2208
 QY 3722 TTCCGAAATGTTATGAAGCAGGATGACTCTGGAGCCCGGAGCTGGAAGAACTTGGA 3781
 Db 2209 TTCCGAAATGTTATGAAGCAGGATGACTCTGGAGCCCGGAGCTGGAAGAACTTGGA 2268
 QY 3782 ATCTGAACATACAGGAGGAGAGGCTTCCAGCACACACAGCCCACTGAGGAGACAA 3841
 Db 2269 ATCTGAACATACAGGAGGAGAGGCTTCCAGCACACACAGCCCACTGAGGAGACAA 2328
 QY 3842 CCCAGAGCTGACAGTGTACACATTTGAAGGCTATGAATGTGAGCCCACTTTCTGAATG 3901
 Db 2329 CCCAGAGCTGACAGTGTACACATTTGAAGGCTATGAATGTGAGCCCACTTTCTGAATG 2388
 QY 3902 TCCTGGAAGCCATTGAGCCAGGTGTAGTGTGCTGGACACGACAAACACAGCCCGACT 3961
 Db 2389 TCCTGGAAGCCATTGAGCCAGGTGTAGTGTGCTGGACACGACAAACACAGCCCGACT 2448
 QY 3962 CCTTTGAGCCTTCTCTAGCTCAATGAACCTGGAGAGACAGCTGTGACAGTGG 4021
 Db 2449 CCTTTGAGCCTTCTCTAGCTCAATGAACCTGGAGAGACAGCTGTGACAGTGG 2508
 QY 4022 TCAAGTGGGCAAGCCCTTCCCTGGCTTCGCGCAACTTACAGTGGAGCAGCAGATGGCTG 4081
 Db 2509 TCAAGTGGGCAAGCCCTTCCCTGGCTTCGCGCAACTTACAGTGGAGCAGCAGATGGCTG 2568
 QY 4082 TCATTGAGTACTCTGTGATGGGCTCATGTGTTGCCATGGCTGGCATCTTCCACCA 4141
 Db 2569 TCATTGAGTACTCTGTGATGGGCTCATGTGTTGCCATGGCTGGCATCTTCCACCA 2628
 QY 4142 ATGTCAACTCCAGGATGCTTACTTCCGCTGATCTGTGTTTCAATGAGTACCGCATGC 4201
 Db 2629 ATGTCAACTCCAGGATGCTTACTTCCGCTGATC1GGTTTCAATGAGTACCGCATGC 2688
 QY 4202 ACAAGTCCCGGATGACAGCAGTGTGTCGATGAGGACCTCTCTCAGAGTGTGGAT 4261
 Db 2689 ACAAGTCCCGGATGACAGCAGTGTGTCGATGAGGACCTCTCTCAGAGTGTGGAT 2748
 QY 4262 GGCTCCCAATCACCCCGCAGGAATTCCTGTGATGAAGCACTGCTACTCTTCAGCATTA 4321
 Db 2749 GGCTCCCAATCACCCCGCAGGAATTCCTGTGATGAAGCACTGCTACTCTTCAGCATTA 2808
 QY 4322 TTCCAGTGGATGGGCTGAAAAATCAAAATTTCTTTGATGAACCTTCGAATGAACATCA 4381
 Db 2809 TTCCAGTGGATGGGCTGAAAAATCAAAATTTCTTTGATGAACCTTCGAATGAACATCA 2868
 QY 4382 AGGAAGTCCGATGATCATTCATTCGATGCAAAAGAAAAATCCACATCTGCTCAAGAGCT 4441
 Db 2869 AGGAAGTCCGATGATCATTCATTCGATGCAAAAGAAAAATCCACATCTGCTCAAGAGCT 2928
 QY 4442 TCTACAGCTCACCAAGCTCTGGACTCCGTGCAGCCTATTGCGAGAGAGCTGCATCAGT 4501
 Db 2929 TCTACAGCTCACCAAGCTCTGGACTCCGTGCAGCCTATTGCGAGAGAGCTGCATCAGT 2988
 QY 4502 TCACCTTTGACCTGTAATCAAGTACACATGGTGGAGCTGGACTTTCCGGAATGATGG 4561
 Db 2989 TCACCTTTGACCTGTAATCAAGTACACATGGTGGAGCTGGACTTTCCGGAATGATGG 3048
 QY 4562 CAGAGATCATCTCTGCAAGTGCCCAAGATCTTCTTGGGAAAGTCAAGCCCATCTATT 4621
 Db 3049 CAGAGATCATCTCTGCAAGTGCCCAAGATCTTCTTGGGAAAGTCAAGCCCATCTATT 3108
 QY 4622 TCCACACCCAGTGAAGCATTTGGAACCCCTATTTCGCCACCCAGCTCATGCCCTTTCA 4681

Db 3109 TCCACACCCAGTGAAGCATTTGAAACCCCTATTTCCACCCAGCTCATGCCCTTTCA 3168
 QY 4682 GATCTCTTCTGCTGTTTATAACTCTGCACCTACTCTCTGCAGTGCCTTGGGGAATTCCT 4741
 Db 3169 GATCTCTTCTGCTGTTTATAACTCTGCACCTACTCTCTGCAGTGCCTTGGGGAATTCCT 3228
 QY 4742 CTATTGATGATACAGTCTGTCATGAACATGTTCTCTGAATCTATTGCTGGCTTTT 4801
 Db 3229 CTATTGATGATACAGTCTGTCATGAACATGTTCTCTGAATCTATTGCTGGCTTTT 3288
 QY 4802 TCTCTTCTCTCCTTCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4861
 Db 3289 TCTCTTCTCTCCTTCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3348
 QY 4862 AGACTTTGCTTCCCATGTTGGCTCTCTATCTGTGTTTGAATGTTGTTGATGCTTTAAA 4921
 Db 3349 AGACTTTGCTTCCCATGTTGGCTCTCTATCTGTGTTTGAATGTTGTTGATGCTTTAAA 3408
 QY 4922 TCTGTGATGATCCTCATATGCCCCAGTGTCAAGTGTGCTGTTTACAGCACTACTCTGT 4981
 Db 3409 TCTGTGATGATCCTCATATGCCCCAGTGTCAAGTGTGCTGTTTACAGCACTACTCTGT 3468
 QY 4982 GCCAGCCACACAAAGCTTTACTTATCTTATGCCACGGAAAGTTTAGAGAGCTAAGATTAT 5041
 Db 3469 GCCAGCCACACAAAGCTTTACTTATCTTATGCCACGGAAAGTTTAGAGAGCTAAGATTAT 3528
 QY 5042 CTGGGGAATCAAAACAAAAAACAAGCAAAAAA 5082
 Db 3529 CTGGGGAATCAAAACAAAAAACAAGCAAAAAA 3569

RESULT 7
 AAS93472/c
 ID AAS93472 standard; cdNA; 3590 BP.
 XX
 AC AAS93472;
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29276.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic; ss.
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 (HYSE-) HYSEQ INC.
 XX
 Drmanac RT, Liu C, Tang YT;
 XX
 WPI: 2001-639362/73.
 DR P-PSDB; ABG29285.
 XX
 New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 Claim 1; SEQ ID No 29276; 103pp; English.
 PS
 The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The

PA (UYN-) UNIVERSITY OF NORTH CAROLINA.
 XX French FS, Wilson EM, Joseph DR, Lubahn DB;
 XX WPI; 1989-324206/44.
 DR P-FSDB; AAP93110.
 XX
 PT DNA encoding androgen receptor protein - useful for transforming
 PT eukaryotic hosts for protein expression and subsequent antibody prodn.
 XX Disclosure; Fig. 5; 41pp; English.
 XX
 CC Complementary DNA sequences derived from the cDNA may be used as probes
 CC to detect the presence of androgen receptor (AR) mRNA in tumour cells, and
 CC to detect AR gene defects using DNA hybridisation assays.
 XX
 SQ Sequence 4180 BP; 1024 A; 1149 C; 1083 G; 924 T; 0 other;

Query Match 48.9%; Score 2486.4; DB 10; Length 4180;
 Best Local Similarity 77.9%; Pred. No. 0;
 Matches 3416; Conservative 0; Mismatches 664; Indels 306; Gaps 23;

QY 758 ATCCCGGGAGCCAGCTGCTGGAGAGCGGGAACGCTCCGGAGAGCCCGAGAGCGAGA 817
 DB 1 ATCCCTAGGAGCCAGCTGCTGGAGAACAG-AGGGTCCGGAGCAAACTGGAGGCTGA 59
 QY 818 GGAGGCGACAGAGGGAAGGAGCCNAGCTAGCCGCTCCAGTGCCTGTACAGNAGCCGAA 877
 DB 60 GAGGGCATCAGAGGGGAAAAA--CTGAGTTAGCCACTCCAGTGCATACAGAGCTTAA 117
 QY 878 -GGACGACACACCCAGCCAGCCGCTCCAGCGACAGACGACGACGACGACGCTTTCGANGCGT 936
 DB 118 GGCACATACACCCAGCCAGCCGCTCCAGCGACAGACGACGACGACGCTTTCGANGCGT 177
 QY 937 TCGAAGCCGCGGCC -GGAGCTGCCCTTTCCTCTTCGCTGAACTTTTAAAGAGCTGCTAA 995
 DB 178 TCGAAGCCGCGGCCAGAGCTGCCCTTTCCTTCGCTGAACTTTTAAAGAGCTGCGGG 237
 QY 996 AGACTCGGAGGAGCAAGAAAGTGGTGGAGTACGAGTACGAGTACGAGTACGAGTACGAGT 1055
 DB 238 AGACTCGGAGGAGCAAGAAAGTGGTGGAGTACGAGTACGAGTACGAGTACGAGTACGAGT 296
 QY 1056 TCTCCACCCGCGCTCCCGCCAGCCCTGCTTCCCGCCCTCCCGCCCTCCCGCCCTCCCGCC 1115
 DB 297 CTCCTACCC -CTACCCCTCTCGGCTCCCGCCCTCCCGCCCTCCCGCCCTCCCGCCCTCC 335
 QY 1116 TGCCTCAGTGGCTACTCTCAGCAACCCCTCACCACCCCTTCCCGCCCTCCCGCCCTCCCGCC 1175
 DB 336 GGACTAGGCAGGCTTC -CTGCGCCAGCC 361
 QY 1176 CGCCCGCGCTCGCCCGAGCGTNCAGNCCGAGTTTGCAGAGAGTAACTCCCTTTGGGTG 1235
 DB 362 CTCCTCCCTACACACAGCTCTGCCGCGAGTTTGCAGAGAGTAACTCCCTTTGGGTG 421
 QY 1236 CGAGCGGCGAGNCTAGCTGCATTTGCAAGAGGCTCTTAGGAG -CAGCGAGCTGGG 1294
 DB 422 AAGCAGAGCAG -CTTGTGTCCTTGAAGGAGGCTTTTGGAGGCCAGAGACTGAGG 480
 QY 1295 AGCGGCTTCAGACTTCAGACGACGACGACGCTGTTAGGCTGACGCGGAGAGAACCCCTC 1354
 DB 481 AGCAACAGCAGCTGGAGAGTCCCTGATTCAGGTT - 516
 QY 1355 TGTTTTCCCGCCACTCTCTCCACT 1412
 DB 517 -CTCCCGCTTCAGCT 565
 QY 1413 AGAGATCAAAAGATGAAAGGAGTCAAGTCTTCAGTAGCCAAAGAAACAAACAAACAA 1472
 DB 566 AGAATGAAAGATGAAAGACAGTGGGGCTTCAGTAGTCGAAGCAAAAC - 617
 QY 1473 AACAAAGCGGAAATGAAAGAAAGATTAATTAAGTCAAGTCTTATTTGACCACTTTC 1532
 DB 618 -AAAGCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 672

QY 1533 AGTGACACTGAATTTGGAGGTGGAGGATTTTGTCTTTTCTTTTAAAGATCTGGGCATC 1592
 DB 673 AGTGGACATTTGACTTTGGAGGACAGAGATTTCTTCCCCCAGTCAAGCTTTGAGCAT 732
 QY 1593 TTTTGAATCACTCCCTCAAGTATTAAAGACAGACTGTGAGCTAGCAGGGCAGATCTTG 1652
 DB 733 CTTTAAATCTGTTCTTCAAGTATTAGGACAACTGTAAACTAGCAGGGCAGATCTCG 792
 QY 1653 TCACCGCTGTCTTCTTCTGACAGACTTTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTG 1712
 DB 793 TCTAGCGGTGCTTCTTCTTACAGGAGACTTTGAGGCTATCTGGCGCT -CCCCC 847
 QY 1713 TCTCTCCCAAGTCTTCTTCTGAGCTTCCCGCAGCTGGCAGCTAGCTAGCTGACGAGCT 1772
 DB 848 CTTCCCTGCAAGTTTCTTCTTCCCTGGAGCTTCCCGCAGCTGGCAGCTAGCTAGCTAGCT 907
 QY 1773 ACCGATCATCAGCCCTGTTGAAGTCTTCTGAGCAAGAGAGAGGGGAGGGGTAAGGG 1832
 DB 908 ACATCAT -CAGTCACTAGACTTCTCAGAGCAAGAGAGAGGAGGAGGAGGTAAGGG 962
 QY 1833 AAGTAGTGAAGATTCAAGCAAGCTCAAGATGGAAGTGCAGTTAGGCTGGAAGGCT 1892
 DB 963 AATTCGGTGAAGCTAGAGACAGCTAAGGATGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 1022
 QY 1893 CTACCTCGCGCGCTCCAGACCTTACCGAGGAGCTTTCCAGAACTCTGTTCAGAGCGT 1952
 DB 1023 CTACCCAGCGCGCTCCAGACCTTCCAGAGCTTCCAGAGCTTCCAGAACTCTGTTCAGAGCGT 1082
 QY 1953 GCGGAGTGTATCCAGAACCCCGCCAGCACCAGCAGCGCGGAGCGCTAGCAGAGCTAGCAG 2012
 DB 1083 GCGGAGCGATCCAGAACCCCGCGCGCAGCAGCTTCCAGAGCTTCCAGAACTCTGTTCAGAGCGT 1142
 QY 2013 CGGCGCAGTTTGTCTGTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2072
 DB 1143 CGGTGCTGTTCAGCAGCGCGCAGGAGCTAGCCCCCGGC - 1183
 QY 2073 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2132
 DB 1184 -GGCGGCGGCG 1193
 QY 2133 GCAGCAGCAGGCTGAGGATGTTCTCCCAAGCCCTCTGAGAGCGCCCAAGCTTACCT 2192
 DB 1194 GCAGCAGCAGCTTCCAGGATGCTCTCTCAAGCCACATCAGAGCAGCAGCAGCAGCTACCT 1253
 QY 2193 GGTCTGTGATGAGCAAGCAACTTTCAGAGCGCAGCTCGGCGCTGGAGT3CCAACCCGA 2252
 DB 1254 GGCCTGGAGGAGCAAGCAGCGCTTCCAGCAGCAGCTAGCAGCTCGAGGCGCCAGCTGA 1313
 QY 2253 GAGAGTGTGCTCCAGAGCTGAGCGCGCTGGCGCCAGCAAGGGCTGCGGAGCA 2312
 DB 1314 GAGCGGTGCTCCCGAGCTGGAGCTGCCAGGCTTCTGCGAGGGGCTGCGGAGCA 1373
 QY 2313 GCTGCCAGCAGCTCCCGAGCAGGATGACTCAGTGCCTCCCAATCCACCTGTCTCTGCTGG 2372
 DB 1374 GCCACCACTCTCCAGATCAGGATGACTCAGCTGCCCATCCACCTGTCTCTACTGG 1433
 QY 2373 CCCCACTTCCCGCTTAAAGCAGCTGCTCCGCTGAGCTTAAAGACATCTTGGCAGGCG 2432
 DB 1434 CCCCACTTCCCGCTTAAAGCAGCTGCTCCGAGAGCTTAAAGACATCTTGGCAGGCG 1493
 QY 2433 CAGCAGCATGCACTCCTT - 2451
 DB 1494 CGCAGCATGCACTTCTTCCAGCAGCAGCAACAGCAACAGCAGCAGCAGCAGCAGCA 1553
 QY 2452 -CAGCAACAGCAGCAGCAGCAGTATCCGAAGGAGCAGCAGCAGGAGAGC 2501
 DB 1554 GCAGCAGCAGCAGCAGCAGCAGGAGTATATCCGAAGCAGCAGCAGCAGCAGCAG 1613
 QY 2502 GAGGAGGCGCTGGGCGCTCCCACTTCTTCCCAAGGAGCAATTTAGGGGAGCTTCGAC 2561
 DB 1614 AAGGAGGCGCTGGGCGCTCCCTCTTCTTCCCAAGGAGTATTTACCTAGGGGCGCAATTCGAC 1673

2562 CATTTCTGACACGCGCAAGGAGTTGTGTAAGCAGTGTGCGTGTCCATGGCCCTGGGTGT 2621
 1674 CATATCTGACAGTGCACAGGAGTTGTGTAAGCAGTGTGTCCTCATGGGTGGGTGT 1733
 2622 GGAGCGTGGAGCAGTCTCAGTCCAGGGAACAGCTTGGGGGATTCATGACGCCCC 2681
 1734 GGAAGCAGTGGAAATCTCTGAGTCCAGGGAGCAGCTTGGGGCGACTGCATGTACCGCTC 1793
 2682 ACTTTTGGGAGTTCCACCGCTGTGCGTCCCACTCTTGTGCCCCATTTGGCCGAATGCAA 2741
 1794 GCTCTGGGAGTTCCACCGCTGTGCGTCCCACTCTTGTGCGCTCTGGCCGAATGCAA 1853
 2742 AGTTCTCTGCTAGACAGACGCGCAGGGAAGACACTGAAGATAGTCTGAGTATTCGCC 2801
 1854 AGTCTTTCCTGGAGCAAGCGCGGCAAGGCACTGAAGAGACTGCTGAGTATTCCTC 1913
 2802 TTTCAAGGAGGTTACACCAAGAGGCTAGAAGCGAGAGGCTAGGCTGCTTGGCAGCGC 2861
 1914 TTTCAAGGAGGTTACCCCAAGAGGTTGNAAGTGAGAGTCTGGGCTGCTTGGCAGCAG 1973
 2862 TGCAGCAGGAGGCTCGGGACACTTGAAGTGGCGTCTACCCCTGCTCTCTACAAGTCCGG 2921
 1974 TGAAGCAGGTAGCTTGGGACACTTGAAGTCCCGTCTCTCTCTCTATTAAGTCTGG 2033
 2922 AGCACTGGAGCAGGAGTCTGCTTACAGAGTCCGAGTCTACAACTTTCCACTGGCTCT 2981
 2034 AGCAGTGAAGCAGCAGCAGTACCAAGATCCGAGTCTACTACAACTTTCCCGTCCGCTCT 2093
 2982 GCGCGAGCGCGCCCTCCCGCGCTCCCACTCCAGCTCCGATCAAGTCTGGAGAA 3041
 2094 GTCCGGCGCGCGCAGCCCGCCCGCTCCCACTCCAGCTCCGATCAAGTCTGGAGAA 2153
 3042 CCCGCTGGAGTACGGCAGCGCTGGCGGCTGGCGGCGCAGTGGCGCTATGGGACCT 3101
 2154 CCCGCTGGAGTACGGCAGCGCTGGCGGCTGGCGGCGCAGTGGCGCTATGGGACCT 2213
 3102 GCGAGCGCTGATGGCGGGGTGAGCGGAGCGGCTTCTGGGTCACTCCCTCAGCCCGCG 3161
 2214 GCGTAGCCTATAGTGGAGGAGTGTAGCGGAGCGGAGTGGATCGGCGCCCGCAGCCCGC 2273
 3162 TTCCTCATCTCTGACACTCTCTTCAAGCGCGGAAGAGGCGAGTGTGATGGACCGTGTGG 3221
 2274 CTCCTTCTCTGCGATCTCTCTTCAAGCTGAGAGGCGCAATATATGGGC ----- 2327
 3222 TGGTGTGGGGGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3281
 2328 -----AGGAGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2348
 3282 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3341
 2349 TAGCCCAAGCGATGCTGGGCGCTGTAGCCCGCTGTAGGCTTACACTCGGCGCCCTCAGGGGCT 2408
 3342 GCGCGGCGGAGGAAGCGACTTACCGCAGCTGATGTGTGGTACCTTGGCGGCGATGGTGAG 3401
 2409 GCGAGCGGAGGAGTCTCTCTGCTCTGAGTGTGGTATCTCTGTTGGAGTGTGAA 2468
 3402 CAGAGTGGCTTATCCAGTCCCACTTGTCTAAAGCGAATGGCCCTGGATGGATAG 3461
 2469 CAGAGTCCCTATCCAGTCCCACTTGTCTAAAGCGAATGGCCCTGGATGGATAG 2528
 3462 CTACTCCGAGCTTACGGGAGCATCGTTTGGAGACTGCCAGGAGCATTGTTTGGCCAT 3521
 2529 CTACTCCGAGCTTATGGGAGCATCGTTTGGAGACTGCCAGGAGCATTGTTTGGCCAT 2588
 3522 TGACTATTACTTTCACCCCGAGAGCCTGCTGATGTGTGGAGATGAAGCTTCTGGGTG 3581
 2589 CGACTATTACTTTCACCCCGAGAGCCTGCTGATGTGTGGAGATGAAGCTTCTGGGTG 2648
 3582 TCAGTATGAGCTCTCAGATGTGGAGCTGCAAGCTCTCTTCAAGAGCGCGCTGAAGG 3641
 2649 TCAGTACGAGCTCTCAGTGTGGAGCTGCAAGCTCTCTTCAAGAGCGCTGCGGAAGG 2708
 3642 GAAACAGAGTACCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3701

2709 GAAACAGAGTATCTATGTGCCAGCAAGATGATGCCAATGATAAAATTTCCGAGGAA 2768
 3702 AATTTGCCATCTTGTGCTCTCGGAATCTTATGAAGCAGGATGACTCTCGGAGCCCG 3761
 2769 AATTTGCCATCTGCTCTCGGAATCTTATGAAGCAGGATGACTCTCGGAGCTCG 2828
 3762 GAAGCTGAAGAACTTGGTAACTTACAGAGGAGGAGAGGCTTCCAGTCCAC 3821
 2829 TAAGCTGAGAACTTGGAACTTCAAACTTACAGGAAGAGGAGAACTTCCAGTCTGG 2888
 3822 CAGCCCCACTGAGGAGACACCCAGAACTGACAGTGTACACATTTGAAGCTTGAATG 3881
 2889 TAGCCCCACTGAGGAGCCATCCAGAAAGTACTGTATCACATTTGAAGGCTTGAATG 2948
 3882 TCAGCCCATCTTCTGAATGTCTGGAAGCATTTGACAGGCTGTAGTGTGTGCTGGACA 3941
 2949 TCAACCTATCTTCTTAACTGCTGGAAGCCATTTGACAGGAGTGTGTGCTGGCGACA 3008
 3942 CGACAACACAGCCGACTCTTTGACGCTTCTCTAGCCTCAATGACTGGGAGA 4001
 3009 TGACAAACACAGCTGATTTCTGCTGCTTATCTAGTCTCAACGAGCTTGGCA 3068
 4002 GAGAGCTTGTACAGTGTGCAAGTGGGCGCAAGCTTGGCTGGCTTCCGCAACTTGCA 4061
 3069 GAGAGCTTGTACATGTGTCAAGTGGGCGCAAGCTTGGCTGGCTTCCGCAACTTGCA 3128
 4062 CGTGACACACAGATGGCTGCTCACTAGTCTCTGATGGGCTCATGTTGTTGGCAT 4121
 3129 TGTGATGACAGATGGGAGTCACTAGTATCTGATGGGCTCATGTTGTTGGCAT 3188
 4122 GGGCTGGCGATCTTCACTCACTCACTAGGATGCTTACTTGGACCTGACCTGGT 4181
 3189 GGGTGGCGGCTTCACTCACTAGGATGCTTACTTGGACCTGACCTGGT 3248
 4182 TTTCAATGAGTACCGCATGCAAGTCCGGATGTACAGCCAGTGTGCCAATGAGGCA 4241
 3249 TTTCAATGAGTACCGCATGCAAGTCCGAATGTACAGCCAGTGTGCCAATGAGGCA 3308
 4242 CTTCTCTCAAGAGTGTGATGCTCAAACTCACTCCAGGAAATTCCTGTGATGAAAGC 4301
 3309 CTTCTCTCAAGAGTGTGATGCTCAAACTCACTCCAGGAAATTCCTGTGATGAAAGC 3368
 4302 ACTGTCTCTTACAGATTTTCCAGTGGATGGGCTGAAATCAAAATCAAAATTTCTTTGATGA 4361
 3369 ACTGTCTCTTACAGATTTTCCAGTGGATGGGCTGAAATCAAAATCAAAATTTCTTTGATGA 3429
 4362 ACTTGAATGAACTACATCAAGAACTTGAATGCTATGATGCAAAAGAAATTTCC 4421
 3429 ACTTGAATGAACTACATCAAGAACTTGAATGCTATGATGCAAAAGAAATTTCC 3488
 4422 CACATCTCTCAAGAGCTTCTACAGCTCCTCAAGCTTCTGAGCTCGTGGACCTAT 4481
 3489 CACATCTCTCAAGAGCTTCTACAGCTCCTCAAGCTTCTGAGCTCGTGGACCTAT 3548
 4482 TGGGAGAGCTGCTGATGCTTCACTTTGACCTGCTTAACTCAAGTCAACATGCTGAGCT 4541
 3549 TGAAGAGAGCTGCTGATGCTTCACTTTGACCTGCTTAACTCAAGTCAACATGCTGAGCT 3608
 4542 GGACTTCCGGAATGATGCGAGATCATCTCTGTCAGTGGCAAGTGGCAAGCTTCTTCTGG 4601
 3609 GGACTTCCGGAATGATGCGAGATCATCTCTGTCAGTGGCAAGTGGCAAGCTTCTTCTGG 3668
 4602 GAAAGTCAAGCCCATCTATTTCCACAGCAGTGAAGCTTGGAAAGCTTATTTCCACCC 4661
 3669 GAAAGTCAAGCCCATCTATTTCCACAGCAGTGAAGCTTGGAAAGCTTATTTCCACCC 3728
 4662 CCAGCTATGCCCCCTTTCAGATGCTTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 4719
 3729 CCAGCTGTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 3787
 4720 GCAGTGCCTT -GGGGAATTTCCCTTATGATGTACAGTGTGCTGATGAAGCTGTTCTGAA 4778

QY	1461	AAAAAACAACAAAAACGCGAAATAAAAAGAAAAAGATAATAACTCAGTCTCTATT	1520
DB	3587	AAAAAACAACAAAAACGCGAAATAAAAAGAAAAAGATAATAACTCAGTCTCTATT	3528
QY	1521	TGCACCTACTTCAGTGGACACTGAATTTGCAAGGTGGAGATTTGTGTTTTCTTTTAA	1580
DB	3527	TACACCTACTTCAATAAACACTAAATTTAAAAAATAAAAAATTTATTTTTCTTTTAA	3468
QY	1581	GATCTGGGCATCTTTTGAATACCTCTCAAGTATTAAAGACAGACACTGTGAGCGTAGCA	1640
DB	3467	AATCTAAACATCTTTTAAATCTACCTTCAAAATATTAAAAACAAACTATAAACCTAACA	3408
QY	1641	GGCGAGATCTTGTCCACCGTGTGCTTCTTCTGCACGAGACTTTGAGGCTGTACAGCGC	1700
DB	3407	AAAAAATCTTATCACCGGTATATCTTCTTCTACAGAAATTTAAACATATCAAAACGC	3348
QY	1701	TTTTTCGTGGTTGTCCTCCGCAAGTTCTTCTCTGAGCTTCCCGCAGGTGGGACGTA	1760
DB	3347	TTTTTACGTAATTAATCTCCGCAATTTCCCTCTTAAACTTCCCGCAATAAACAACTA	3288
QY	1761	GCTGCAGGACACTACCGCATCATCACAGCTGTGAACTCTTCTGACACAGAAGGGAG	1820
DB	3287	ACTACAACGACTACCGCATCATCACAGCTATTAAACTCTTTTAAACAAAAAATAAAA	3228
QY	1821	GCGGGTATAGGGAAGTAGTGGGAAGATTACGCCAAGCTCAAGGATGGAAGTCAGTTAGG	1880
DB	3227	ACGAATATAAAAAATAATAAAAAATTCACCAAACTCAAAATAAAAAACAATATAA	3168
QY	1881	GCTGGGAAGGTCTACCTCTGGCGCGCTGCCAAGACTACCGAGGAGCTTTCAGAACTC	1940
DB	3167	ACTAAAAAAATCTACCTCTCGACCCGCTCCAAAACCTTACGAAAAACTTCCAAATCT	3108
QY	1941	GTTCAGAGCTGCGGGAAGTGCATCAGAACCGGGGCCAGGCAACCCAGAGCGCGAG	2000
DB	3107	ATTCAAAAAGTACGGAAATAATCCAAACCCGNAACCCCAACAAAAACCGCGAA	3048
QY	2001	CGCAGCACTCCCGGCGCAGTTTCTGCTGTGTGACGAGCAGCAGCAGCAGCAGCAGCA	2060
DB	3047	CGCAACACTTCGCGAGCGCAATTTACTACTACTACACACAAACAAACAAACAAAC	3001
QY	2061	GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	2120
DB	3000	-----AC	2952
QY	2121	CAGCAGCAGCAGCAGCAGCAGGTTGAGATGGTTCTCCCAAGCCCATCGTAGAGCCG	2180
DB	2951	CAAAACAACAACAACAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2892
QY	2181	CACAGGTACTGTGCTGAGTGGAGAACAGCAACTTCACAGCGCGAGTCGGCCCTGGA	2240
DB	2891	CACAACCTACTCTAATCTTAATAAANAACACACACTTCACACGCAATCGACCCATAA	2832
QY	2241	GTGCCACCCGAGAGGTTTGGCTCCAGAGCCTGGAGCGCGCTGGCGCGCAGCAGG	2300
DB	2831	ATACACCCCGAAAAAATTTACGTCCTCAAAAGCTTAAACCCGCGCTAACCGCAACAAA	2772
QY	2301	GCTGCCGACAGTGCACACACTCCGACAGAGTAGCTCAGCTGCCCCATCCAGCTT	2360
DB	2771	ACTACCGCAACACTCCAACACTCCGAACGAAATAACTCACTACCCATCCAGCTT	2712
QY	2361	GTCCCTGTGGGCCCCACTTTCCCGGGCTTAAGACGTGCTCCGCTGACCTTTAAGACAT	2420
DB	2711	ATCCCTACTAAACCCCACTTTCCCGACTTAAACAACACTTCGCTTAACCTTAAAAACAT	2652
QY	2421	CCTGACGAGCGCAGCACTGCACTCTTTCAGCAACAGCAGGAGGAGCAGTAGTATCCGA	2480
DB	2651	CGTAACAGAAACCAACACCACTACAACTCCCTTCAACAAACAAACAAACAAATATCCGA	2592
QY	2481	AGGCAGCAGCAGCGGGAGCGAGGAGGCGCTTCGGGGCTCCCACTTCTCTCCAGGACAA	2540
DB	2591	AAACAACAACAACGAANAACGAANAACCTCGAATAACTCCCAACTCTCTCCCAAAAAA	2532
QY	2541	TTACTTAGGGGCACTTCGACCAATTTCTTGACAAAGCCAGGAGTTCGTAGGCACTGTC	2600

Db	2531	TTACTTAAAAACACTTCGACCAATTTCTAACACGCCCAAAAAATATATATAAAACAATATC	2472
Qy	2601	GGTGTCATGGCCCTGGGTGTGGAGCGTTGGAGCATCTGAGTCCAGGGGAACAGCTTCG	2660
Db	2471	GATATCCATAAACCTTAATATAAAAACGTTATAAACATCTTAATCCAAAAAACACACTTCG	2412
Qy	2661	GGGGATTGCATGTACGCCCCACATTTTGGGAGTTTCCACCGCTGTGCGTCCCATCTCCTTG	2720
Db	2411	AAAAAATATACATATACGCCCCACATTTTAAAAAATTTCCACCGCTATACGTCCTCCTTA	2352
Qy	2721	TGCCCCATTTGGCCGAATCAAAAGTTCTCTCTAGACGACGAGCGAGCGACGAGCACTGA	2780
Db	2351	TACCCCAITTAACCGAATACAAAAATCTCTACTAAACGACACGCAAAACAAACACTAA	2292
Qy	2781	AGATACTGCTGAGTATTCCCTTTTCAAGGGAGGTTTACACCAAGGCTAGAGGCGAGAG	2840
Db	2291	AAAAATCTACTAAATATTCCCTTTTCAAAAAAATTTACACCAAAAAAATTA AAAACGAA	2232
Qy	2841	CCTAGGCTGCTTGGCAGCGCTGCAGCAGGAGCTCCGGGACACTTGAACCTGCGTCTAC	2900
Db	2231	CCTAAACTACTCTACACAGCTACACAAAAAACTCCGAAACACTTAAACTACCTCTAC	2172
Qy	2901	CCTGTCCTCTTACAAGTCGGAGCACTGGACGAGGACGCTGGTACACAGATCGCGACTA	2960
Db	2171	CCTATCTCTCTACAATCGAAACACTTAACGAAAACTACGTACCAAAATCGGACTA	2112
Qy	2961	CTACAACCTTCCACTGGCTCTGGCGGACGCGCGCCCTCGCGCGCTCCCATCTCCCA	3020
Db	2111	CTACAACCTTTCACATAACTCTAACCGAACCGCGCCCTCGCGCGCTCCCATCTCCCA	2052
Qy	3021	CGCTCGCATCAAGCTGGAGAACCCGTGGACTACGGCAGCGCTGGGCGGCTGCGGCGC	3080
Db	2051	CGCTCGCATCAAACTAAAAAACCCGTAACTACGACAACGGCTTAACGACTACGACGAC	1992
Qy	3081	GCAGTCCGCTATGGGACCTGCGGAGCCTGCTATGCGGGGTGCAGCGGGACCGGTTTC	3140
Db	1991	GCAATACCGCTATAAAAACCTTACGAACCTACATAACGCGAATACACGAAACCGGATTC	1932
Qy	3141	TGGGTCACCTCATAGCCGCGCTTCCTCATCTCTGGCACACTCTCTTCACGCGGAGAAGG	3200
Db	1931	TAAATCACCCTCAACCGCGCTTCCTCATCTTCAACACACTCTCTTCAACACCGGAAAA	1872
Qy	3201	CCAGTTGTATGGACGTGTGGTGGTGGTGGGT-----GGTGGCGGCGCGCGCGG	3251
Db	1871	CCAATTATATAAACCGCTATAATAATAATAATAATAATAATAACGACGACGACGACGA	1812
Qy	3252	CGGAGCGGAGCTGTAGCC	3311
Db	1811	CGAGGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	1752
Qy	3312	CTACGGCTACACTCGGCGCCCTCAGGGGCTGGCGGCGAGGAAAGCACTTCAACCGCAC	3371
Db	1751	CTACGACTACACTCGACCCCTCAAAACCTTAACGAACCAAAAAAGCACTTCAACCGCAC	1692
Qy	3372	TGATGTGTGTACCTCGCGGCACTGGTGACAGATGCGCCTATCCAGTCCCACTTGTGT	3431
Db	1691	TAAATATATAATACCTTAACGACATATAACAAAAATACCTATCCCAATCCCACTTAT	1632
Qy	3432	CAAAAGCGAAATGGGCGCCCTGGATGGATGACTACTCGGACCTTACGGGACATGGTTT	3491
Db	1631	CAAAACGAAATAAACCCCTAAATAATAATAATAATAATACTCTCGGAACCTTACGAA	1572
Qy	3492	GGAGACTGCCAGGACCATGTTTGGCCCATGTACTATTCTTTCCACCCCGAAGACCTG	3551
Db	1571	AAAAAATACCAAAACCATATTTTACCCATTAACATATTACTTTCCACCCCGA AAAACCTA	1512
Qy	3552	CCTGATCTGTGGAGATGAAGCTTCTGGGTGTCACTATGAGCTCTACATCTGGGAGCTG	3611
Db	1511	CCTAATCTATAAAATAAACTTCTAATATCACTATAAAACCTCTACATATAAAAACTA	1452
Qy	3612	CAAGGCTCTTCTTCAAAAGACCGCTGAAGGGAAACAGAAGTACTGTGTGGCGCACGAAA	3671

Db 1451 CAAATCTCTTCCTCAAAAAACCGCTAAAAAAACAAAAACAAAAATACCTATACGCCAACAAAA 1392
QY 3672 TGATTCGACATATTGATAAATTCGGAAGGAAAAATGTCCATCTTGTCTTCGGAATG 3731
Db 1391 TAATACACTATTATAATTCGGAAGGAAAAATATCCATCTTATGCTTCGGAATA 1332
QY 3732 TTATGAAGCAGGATGACTCTGGGAGCCGGAAGCTGAAGAACTTGGTAATCTGAAACT 3791
Db 1331 TTATAAACAATAAATACTTAAACCCGGAACCTAAACAACTTAAATATCTAAACT 1272
QY 3792 ACAGAGAGAGAGAGGCTTCAGCACCACAGCCGCTACTGAGGAGACACCCAGAGCT 3851
Db 1271 ACAAAAAAATAAATACTTCAACACACCAACCCCACTAAAAAACAACCCAAAACT 1212
QY 3852 GACAGTGTACACATTTGAAGCTATGAATGTACGCCATCTTCTGAATGTCCTGGAGC 3911
Db 1211 AACATATCACACATTTAAACTATAATATCAACCCATCTTCTAAATATCTTAAAC 1152
QY 3912 CATTGAGCCAGGTGTAGTGTGTGGACAGCAACAACCCGACCTCTTTGAGC 3971
Db 1151 CATTAAACCAATATAATATTAATTAACACAGCAACAACCCGACCTCTTTTACAC 1092
QY 3972 CTGCTCTCTAGCTCAATGAAGCTTGAATGTACGCCATCTTCTGAATGTCCTGGAGC 4031
Db 1091 CTGCTCTCTAGCTCAATGAAGCTTGAATGTACGCCATCTTCTGAATGTCCTGGAGC 1032
QY 4032 CAAGCCCTGCTGCTTCCGCACTTACAGTGGAGCAGCAGATGGCTGTCTTCAGTA 4091
Db 1031 CAAACCTTCACTAACTTCGCACTTACAGTAAACGCAACCAATTAATCAATCAATA 972
QY 4092 CTCCTGGATGGGCTCATGGTGTGGCATGGGCTGGCATGCTTCCCAATGTCAACTC 4151
Db 971 CTCCTAAATAAATACTATAATTTACCAATAAAGTAACTGATCTTCCCAATTAACA 912
QY 4152 CAGGATCTTACTTCCGCTGATCTGTTTCAATGAGTACGGATGCGATGCGACAGTCCCG 4211
Db 911 CAAATACTCTACTTCCGCTTATCTAATTTCAATAAATGATGATGATGATGATGATG 852
QY 4212 GATGTACAGCAGTGTGTCGAATGAGGCACCTCTCTCAAGAGTTGGATGGCTCCAAAT 4271
Db 851 AATATCAACCAATATATCGAATAAACAACCTCTCTCAAAATTTAAATAAATCAAT 792
QY 4272 CACCCCGCAGGATCTCTGTCATGAAGCAGTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4331
Db 791 CACCCCGCAGGATCTCTGTCATGAAGCAGTCTTCTTCTTCTTCTTCTTCTTCTTCT 732
QY 4332 TGGCTGAAATAAATAAATTTCTTGAATGAGTTCGAATGAATGAGTTCGAATGAGTTCGA 4391
Db 731 TAAACTAAAAATAAATAAATTTCTTAAATAAATGAGTTCGAATGAGTTCGAATGAG 672
QY 4392 TCGTATCATGTCATGAAGGAAAAATCCGATCTCTGTCGATGATGATGATGATGATGAT 4451
Db 671 TCGTATCATGTCATGAAGGAAAAATCCGATCTCTGTCGATGATGATGATGATGATGAT 612
QY 4452 CACCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4511
Db 611 CACCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552
QY 4512 CTTGCTTAATCAAGTACAGTGTGAGCTGTGAGCTTTCCGGAATGATGCGAGATCAT 4571
Db 551 CTTGCTTAATCAAGTACAGTGTGAGCTGTGAGCTTTCCGGAATGATGCGAGATCAT 492
QY 4572 CTTGCTGCAAGTCCCAAGATCTTCTTCTGGGAAGTCAAGCCCATCTATTTCCACACCA 4631
Db 491 CTTGCTGCAAGTCCCAAGATCTTCTTCTGGGAAGTCAAGCCCATCTATTTCCACACCA 432
QY 4632 GTGAAGCATTTGGAACCTTATTTCCGCCAGCTCATGCTGCTGCTGCTGCTGCTGCTGCT 4691
Db 431 ATAAAAATTAATAAATTTTCCGCCAGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 372
QY 4692 GCTGTTAATCAAGTACAGTGTGAGCTGTGAGCTTTCCGGAATGATGCGAGATCAT 4751
Db 371 ACCTATTATAAATCAAGTACAGTGTGAGCTGTGAGCTTTCCGGAATGATGCGAGATCAT 312

QY 4752 ACAGTGTGTCATGAACATGTCCTGAATCTTATTTCTGGCTTTTCTTCTTCTTCTC 4811
Db 311 ACAATCTATCATAA-----AATCTATTACTAACTTTTCTTCTTCTTCTC 263
QY 4812 TCCTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4871
Db 262 TCCTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 203
QY 4872 TCCTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4931
Db 202 TCCTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 143
QY 4932 TCCTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4991
Db 142 TCCTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 83
QY 4992 CAAACCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5051
Db 82 CAAACCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1111
QY 5052 CAAACCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5062
Db 22 CAAACCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 12

RESULT 11
AAQ12002
ID AAQ12002 standard; DNA; 3217 BP.
AC AAQ12002;
DT 20-AUG-1991 (first entry)
DE Full-length rat androgen receptor coding sequence.
KW rAR; DNA-binding protein; steroid hormone; ss.
OS Rattus rattus.
FH Key
FT CDS 33..2741
FT /*tag= a
FT /product= full-length (902 amino acids) rAR
FT /note= "includes 733 residue shorter rAR protein"
XX WO9107423-A.
XX 30-MAY-1991.
XX 19-OCT-1990; 90WO-US06015.
XX 17-NOV-1989; 89US-0438775.
XX (ARCH-) ARCH DEV CORP.
XX Liao S, Chang C;
XX WPI: 1991-178048/24.
XX P-PSDB; AAR12224.
XX Androgen receptor and TR2 DNA binding proteins - DNA sequences
XX and antibodies for detection and quantification methods
XX Claim 4; Fig 3; 79pp; English.
XX This sequence was isolated by screening a rat ventral prostate
XX lambda gt10 library in E.coli Y1090. Initial screening
XX was with probes designed for homology to nucleotide sequences in the
XX DNA-binding domain of known steroid receptors. Positive clones were
XX then screened with 24mer probes specific for the various steroid
XX receptors to eliminate those which coded for known receptors. Any
XX remaining clones were analysed by restriction mapping and were

Db 1924 CTGGTAGCCCACTGAGGACCCATCCAGAGAGTACTGTATCACACATGAAGGCTATG 1983
QY 3878 AATGTCAGCCATCTTCTGAATCTCTGGAAGCCATTTGAGCCAGGTGTAGTGTGCTG 3937
Db 1984 AATGTCACCCATCTTCTTAATGTCCTGGAAGCCATTTGAGCCAGGAGTGTGTGCGC 2043
QY 3938 GACAGGACAACACAGCCGACCTCCCTTTCAGAGCTTGTCTCTAGCCCTCAATGAAGTGG 3997
Db 2044 GACATGACAACAACAGCCGATTCCTTTCCTGCTGCTTGTATCTAGCTCAAGAGAGCTG 2103
QY 3998 GAGAGAGACGCTTGTACACAGTGGTCAAGTGGGCGAAGCCCTTGCCTGCTCCGCAACT 4057
Db 2104 GAGAGAGACGCTTGTACATGCTGGTCAAGTGGGCGAAGCCCTTGCCTGCTCCGCAACT 2163
QY 4058 TACAGCTGGAGACGAGATGGCTGTCAATTCAGTACTCTCTGGATGGGCTCATGGTGTG 4117
Db 2164 TGCATGTGGATGACAGATGCGAGTCATTGAGTATTCCTTGGATGGGAGCTGATGGTATTG 2223
QY 4118 CCATGGGCTGGGATCTTCCACCAATGTCACATCCAGGATGCTTACTTTCGCCCTGATC 4177
Db 2224 CCATGGGCTGGGATGCTTCCACTAATGTCAGTCTAGGATGCTTACTTTCGCCCTGATC 2283
QY 4178 TGGTTTTCAATGAGTACCGCATCCAGATCCAGATCCAGGATGCTTACTTTCGCCCTGATC 4237
Db 2284 TGGTTTTCAATGAGTATCCGATCCAGATCCAGATCCAGGATGCTTACTTTCGCCCTGATC 2343
QY 4238 GGCACCTCTCTCAAGAGTTTGGATGCTCCAAATCACCCCGAGGATTCCTGTGCATGA 4297
Db 2344 GGCACCTCTCTCAAGAGTTTGGATGCTCCAGATCCAGGATTCCTGTGCATGA 2403
QY 4298 AAGCACTGCTACTCTTCCAGCATTTATCCAGTGGGCTGAGGCTGAAATCAAAATCTCTTTG 4357
Db 2404 AAGCACTGCTACTCTTCCAGCATTTATCCAGTGGGCTGAGGCTGAAATCAAAATCTCTTTG 2463
QY 4358 ATGAATTCGATGAATCACTCAAGGACTCGATCGTATCATTCGATCGCAAAAGAAAA 4417
Db 2464 ATGAATTCGATGAATCACTCAAGGACTCGATCGTATCATTCGATCGCAAAAGAAAA 2523
QY 4418 ATCCACATCTCTCAAGAGCTTCTACAGCTTCTACAGCTCACCAGCTCTGAGCTCCGTCAGC 4477
Db 2524 ATCCACATCTCTCAAGAGCTTCTACAGCTTCTACAGCTCACCAGCTCTGAGCTCCGTCAGC 2583
QY 4478 CTATTCGGAGAGCTGTCATCAGTTTCACTTTTGCCTCTAATCAAGTCACACATGCTGA 4537
Db 2584 CTATTCGAAGAGAGCTGCAACAGTTTCACTTTTGCCTCTAATCAAGTCACATGCTGA 2643
QY 4538 GCGTGACTTTCGGAATGATGGCAGAGATCATCTCTGTGCAAGTCCCAAGATCCTTT 4597
Db 2644 GCGTGACTTTCGGAATGATGGCAGAGATCATCTCTGTGCAAGTCCCAAGATCCTTT 2703
QY 4598 CTGGGAAGTCAAGCCCATCTATTTCCACACCCAGTGAAGCATTTGGAACCCCTATTTC 4657
Db 2704 CTGGGAAGTCAAGCCCATCTATTTCCACACACAGTGAAGCATTTGGAACCCCTATTTC 2763
QY 4658 CACCCAGCTCATGCCCTTTCAGATGCTTCTGCTGTCTGCAAGTCCCAAGATCCTTT 4715
Db 2764 AAACCCACTTGT-TCCCTTTTCAGATGCTTCTGCTGTATATAACTCTGCACACTT 2822
QY 4716 CTCTCAGTGGCTT-GGGGAATTTCCCTATTGATGATGATGATGATGATGATGATGATG 4774
Db 2823 CTCTCAGTGGCTTGGGGGAATTTCCCTACTGATGATGATGATGATGATGATGATGATG 2882
QY 4775 TGAATTCATTTGCTGGGCTTTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4834
Db 2883 CAGTTCATTTCTGGGCTTTT-----CCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2930
QY 4835 CCTATCAACCTCCCATGGACCTTCCAGACTTTCCTTCCATGCTGCTGCTGCTGCTGCTGCTG 4894
Db 2931 COTCTTTTACCTCCCATGGACATTTTGAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2989
QY 4895 TTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4954
Db 2990 TTTTGAGTTTGTGATTTCTTCAAGTCTGTGATGATCTTCTTGTGGCCCGAGTGTCAA- 3048

QY 4955 TTGTGCTTGTTCACACAC--TACTCTGTCCAGCCACACAAACGTTTACTTATCTTATG 5012
Db 3049 CTGTGCTTGTTCACACAC--TACTCTGTCCAGCCACACAAACGTTTACTTATCTTATG 3108
QY 5013 CCATGGCAAGTTTACAGAGCTAAGATATCTGGGAAATCAAAACAAACAAACAAAC 5072
Db 3109 CCATGGCAAGTTTACAGAGCTAAGATATCTGGGAAATCAAAACAAACAAACAAAC 3168
QY 5073 AAAAAA 5082
Db 3169 AAAAAA 3178

RESULT 12
AAN91578
ID AAN91578 standard; DNA; 3217 BP.
XX AAN91578;
AC AAN91578;
XX 28-FEB-1990 (first entry)
XX Rat androgen receptor DNA clone.
XX Androgen receptor; TR2 polypeptide;
XX Rat.
XX Key Location/Qualifiers
FT CDS 33..2739
FT /tag= a
FT /product=98 kD polypeptide
FT CDS 541..2739
FT /tag= b
FT /product=79 kD polypeptide
XX W08909223-A.
XX 05-OCT-1989.
XX 24-MAR-1989; 89WO-US01238.
XX 30-MAR-1988; 88US-0176107.
XX (ARCH-) ARCH DEVELOPMENT CORP.
XX Liao S, Chang C;
XX WPI; 1989-309501/42.
XX P-PSDB; AAP91006.
XX New DNA encoding new androgen receptor and TR2 polypeptide(s) - able to bind DNA, and derived antibodies, useful for receptor assay and purification.
XX Claim 8; fig. 3; 60pp; English.
XX The sequence is used to express the corresp. peptides and for hybridisation assays of RNA and DNA encoding androgen receptors.
XX The 98 kD product starts at the first Met codon; the 79 kD product starts from the second.
XX Sequence 3217 BP; 776 A; 874 C; 837 G; 730 T; 0 other;

Query Match 43.0%; Score 2185.8; DB 10; Length 3217;
Best Local Similarity 82.1%; Pred. No. 0;
Matches 2715; Conservative 0; Mismatches 407; Indels 185; Gaps 10;
QY 1832 GAAGTGTGAGATTCAGCCAGCTCAAGGATGAGTCAAGTGTAGGCTGGAAGGG 1891
Db 1 GAATTCGGTGAAGCTAGAGACAAGCTAAAGGATGAGTGTAGGCTGGAAGGG 60
QY 1892 TCTACCTCGGCCGCTCCAGACCTACCGAGGACTTCCAGATCTGTTCAGAGCG 1951

[illegible][illegible]

CC	present sequence is one of the chemically pre-treated reference DNA samples of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 3715 BP; 841 A; 149 C; 1001 G; 1724 T; 0 other;
	Query Match 42.4%; Score 2154; DB 24; Length 3715; Best Local Similarity 74.5%; Pred. No. 0; Matches 2782; Conservative 0; Mismatches 905; Indels 45; Gaps
QY	1340 GCGGAGAACCTCTGTGTTTCCCCACACTCTCTCACCCTCTCTCTGCCCTTCGCCACCC 1399
DB	9 GCGGAGAGAATTTTGTTTTTTTTTTTATTTTTTTTTTTTTGGTTTTTTTATTATTT 68
QY	1400 CGAGTGCGGAGCCAGAGATCAAAAGATGAAGAGGAGCATCAGTCTTCAGTAGCCAAAAA 1459
DB	69 CGAGTGCGGAG-TAGAGATTAAGAGATGAAGAGGAGTAGTTAGTGTTTTASTAGTTAAAAA 127
QY	1460 CAARACAACAACAACAACAACGCCGAAATAAAGAAAAGATAATAACTCAGTCTCTAT 1519
DB	128 TAAATTAATTAATAATAAANAAGTCGAATTAAGNAAGATAATAATTTAGTTTTAT 187
QY	1520 TTGCACCTACTTCAGTGGACACTGAATTTGGAGGTGGAGGATTTTGTTTTTCTTTTA 1579
DB	188 TTGTATTATTATTTAGTGGATTTGAATTTGGAGGTGGAGGATTTTGTTTTTTTTTTA 247
QY	1580 AGATCTGGGATCTTTTGAATCTACCTTCAAGTATTAAAGACACAGACTGTGAGCTAGC 1639
DB	248 AGATTGGGTATTTTTCGAATTTATTTTAAAGTATTAAGACATAGATTGTGAGTTAGT 307
QY	1640 AGGCGACATTTGCCACCGTGTCTCTCTGACAGAGACTTTGAGCTGTCAGACGC 1699
DB	308 AGGCTAGATTTTGTTTATCGTGTCTTTTGTACGAGATTTTACGTTGTTTAGAGCG 367
QY	1700 CTMTTTCGGTGTCTCCCGCAAGTTTCTCTCTGAGACTTCCGCAAGTGGGCAGCT 1759
DB	368 TTTTTTGGTGGTGTGTTTTCGTAAGTTTTTTTTTGGAGTTTTCGTAGGTGGGTAGTT 427
QY	1760 ACTGCAGCGACTACCGCATCATCACAGCCTGTTGAAGTCTTCTGACGAAGAGAGGGA 1819
DB	428 AGTTGTAGCGATATTCGTATTATTATAGTTTGTGTAATTTTTTGTAGTAAGAAGAGGGA 487
QY	1820 GCGGGGTAAAGGAGTAGGTGGAAGATTCAGCCAAGCTCAAGGATGGAAGTGCAGTTAG 1879
DB	488 GCGGGGTAAAGGAGTAGGTGGAAGATTTAGTTAAAGTTAAGGATGGAAGTGTAGTTAG 547
QY	1880 GCCTGGGAAGGCTTACCTCTGGCCGCGTCCAAAGACCTTACCAGAGAGCTTTCCAGATC 1933
DB	548 GGTTGGGAAGGCTTATTTCGTCGTCGTTTAAAGATTATCAGGAGATTTTTTGAATT 607
QY	1940 TGCTCCAGAGCGTGCAGAGTGTATCCAGAACCCGGGCCACCGACCCAGAGCGCGGA 1999
DB	608 TGTTTTAGCGCTGCGCAAGTGAATTAAGAAATTCGGGTTTTAGGTATTTAGAGTCCGCA 667
QY	2000 GCGCAGCACCTCCGGGCCAGTTTGTGCTGCTGACAGCAGCAGCAGCAGCAGCAGC 2051
DB	668 GCGTAGTATTTTTCGGCCTTAGTTGTGTGTGTT-----GT 703
QY	2060 ASCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2111
DB	704 AGT 763
QY	2120 CCAGCGCAGCAGCAGCAGCAGCGGTGAGGATGTTTTCGCCAAGCCCATCTGAGAGCC 217
DB	764 TTAGGTAGTAGTAGTAGTAGTAGGTGAGTGTGTTTTTAAAGTTATTCGTAGAGGTT 823
QY	2180 CCACAGGCTACCTGGTCTCGGATGAGGAACAGCAACCTTCACAGCCGACGTCGCCCTG 223
DB	824 TTATAGGTTATTTGGTTTGGATAGGAATAGTAATTTTATAGTCGTAGCTGTTTTGG 883

Db	3044	ATCGTATTATTGTATGTAAGAAAGAAAAATTTTATATTGTTGTTTAAAGACGCTTTTTATTAGT	3103
QY	4451	TCACCAAGCTCCTGGACTCCGCTGACGCTATTCCGAGAGAGCTGCATCAGTTCACATTTCG	4510
Db	3104	TTATTAAAGTTTTGGATTTTCGTGTAGTTATTCCGAGAGAGTTGTATTAGTTTATATTTCG	3163
QY	4511	ACCTGCTAATCAAGTCACACAGGTGAGCTGGACTTCCGGAAATGATGGCAGAGATCA	4570
Db	3164	ATTTCCTTAATTAAGTTATATATGCTGAGCGTGATTTTTCGAAATGATGGTAGAGATTA	3223
QY	4571	TCCTCTGCCAAGTGCCCAAGATCCTTTCTGGGAAGTCAGGCCATCTATTTCACACACC	4630
Db	3224	TTTTTGTCTAGTGTCTTAAGATTTTTTTGGGAAAGTTAAGTTTAAATTTATTTTATATTT	3283
QY	4631	AGTGAAGCATTTGGAAACCCATTTCCTCCCAACCCAGCTCATGCCCTTTTCAGATGCTTTC	4690
Db	3284	AGTGAAGTATTGGAAATTTTATTATTTTTTATTATTAGTTTATGTTTTTTTAAAGTTTTT	3343
QY	4691	TGCCCTGTTATAAACHCTGCACCTACTCTCTGCACTGCCCTGGGGAATTTCCCTATTGATG	4750
Db	3344	TGTTTGTGTAAATTTTGATATTTTTTTGTGTAGTGTGTTTAAATTTTTTTTATTATGATG	3403
QY	4751	TACAGCTGTCATGAACATGTCCTGAAATCTATTGTCTGGCTTTTTTTTCTCTTTCT	4810
Db	3404	TATAGTTTGTATG- -----GAAATTTATTGTTGGGTTTTTTTTTTTTTTTTTT	3452
QY	4811	CTCCTCTCTTTTCTCTCTCCCTCACTCACTCAACCCCTCCCATGGACCTTCAGACATTTCG	4870
Db	3453	TTTTTTTTTTTTTTTTTTTTTTTTTTTATTAATTTTTTTTATGTTATTTTAGATTGTT	3512
QY	4871	TTCCCATGTGGCTCTATCTGTGTTTTCAATGGTGTGTATGCTTTAAATCTGCTGATG	4930
Db	3513	TTTTTTATTGTGGTTTTATTGTGTTTTCAATGGTGTGTATGTTTTTAAATTTGTGATG	3572
QY	4931	ATCCTCATATGGCCAGTGTCAAGTTGCTGTGTTTACAGCACACTCTGTGCCAGCCAC	4990
Db	3573	ATTTTATATGGTTAGTGTAGTGTGTTTGTGTTATAGTATTATTTTGTGTTAGTTAT	3632
QY	4991	ACAAACGTTTACTTATCTTATGTCACGGGAAGTTTAGAGACCTAAGATTTATCTGGGGAAA	5050
Db	3633	ATAACGTTTATTATTTTATGTTACGGGAAGTTTAGAGATTTAAGATTTATTGGGGAAA	3692
QY	5051	TCAAAACAAAAA 5062	
Db	3693	TTAAAAATAAAA 3704	

XX
XX

(EPTG-) EPIGENOMICS AG.

Qlek A. Piepenbrock C. Berlin K:

WPI: 2002-147896/19.

Oligonucleotide for diagnosis and therapy of diseases associated with signal transduction e.g. cancer, comprises chemically modified genomic sequences of genes associated with signal transduction -

claim 1: SEQ ID No 387; 24pp; English.

The present invention relates to chemically modified DNA sequences of signal transduction associated genes. The DNA sequences are chemically modified using a solution of bisulphite, hydrogen sulphite or disulphite. Also disclosed are oligonucleotides and/or pNA oligomers for detecting the cytosine methylation state (CpG islands) of these genes, and a method for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with signal transduction. The genomic DNA can be obtained from cells or cellular components which contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as tissue from eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver histologic object slides, and all their possible combinations. The sequences of the invention are useful for the diagnosis and therapy of diseases associated with signal transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA sequences of different genes associated with signal transduction, or their complementary sequences.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 3715 BP: 841 A: 149 C: 1001 G: 1724 T: 0 other;

Very Match	42.4%;	Score 2154;	DB 24;	Length 3715;
1st Local Similarity	74.5%;	Pred. No. 0;		
Matches 2782;	Conservative	0;	Mismatches 905;	Indels 45;
			Gaps	4;

1340 GCGGAGAGAACCCCTCTGTTTTCCTCCCACTCTCTCTCCTCCTCCCTGCTTCCCCACCC 1399

0 CCCCCCACACATTTTCTTCTTTAATTTTTTTGTTTATTI 68

[illegible]

107

09 CCGATGCCCCUUU
TACAGCCTTTCTTCCTCCCTGGT

I460 CAAAACAAACAATAAAGGCGAAATATAAGGAAGAATATATATCTTTC

128 TAAATAAATAAAAAATAAAAAAGTCGAAATAAAAAGAAAAGATAAATTAAGTTTAT 187

1520 TTGCACCTACTTCAGTGGACACTGAATTGGAAGTGGAGGATTTGTTTCTTTTCTTTA 1579

188 TTGTATTTATTAGTGCATATTGAATTGGAAGGTGGAGGATTTTGTTTTTT 247

1580 AGATCTGGGCATCTTTTGAATCTACCCCTCAAGTATTAGAGACAGACTGTGAGCCTAGC 1639

248 ACCTTTCCCTATTTTAAAGTATTAGAGATAGATTGTGAGTTTAGT 307

1699

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

1700 CTTTGTGCGTGGTTCCTCCCGCAAGTTCTCTCTGGAGCTTCCCGCAGGTCGGCAGGCT 1750

368 TTTTTCGCGTGGTTGTTTCGTAAGTTTTTTTTTTGGAGTTTTTCGTAGGTGGTAGTT 427

1760 AGCTGCAGCGACTACCGCATCATCAGCCCTGTTGAACTCTTCTGAGCAAGAGAGGGGA 1819

428 AGTTGTAGCGATTATCGTATTATATAGTTGTTGAATTTTGTAGTAAGAGAGGGCA 487

1820 GCGGGGTAAGGGAAGTAGGTGGAAGATTCAGCCAAGCTCAAGGATGGAAGTGCAGTTAG 1879

100

[illegible]

Search completed: May 8, 2003, 15:16:10
Job time : 1176 secs

